



**HAL**  
open science

# Sub-seasonal meteorological predictions for the European energy sector : Quantitative assessment, improvement, and application

Naveen Goutham

► **To cite this version:**

Naveen Goutham. Sub-seasonal meteorological predictions for the European energy sector : Quantitative assessment, improvement, and application. Meteorology. Institut Polytechnique de Paris, 2022. English. NNT : 2022IPPAX125 . tel-04224018

**HAL Id: tel-04224018**

**<https://theses.hal.science/tel-04224018v1>**

Submitted on 1 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NNT : 2022IPPAX125

Thèse de doctorat



# Sub-seasonal meteorological predictions for the European energy sector: quantitative assessment, improvement, and application

Thèse de doctorat de l'Institut Polytechnique de Paris  
préparée à l'École Polytechnique

École doctorale n°626 École doctorale de l'Institut Polytechnique de Paris (EDIPP)  
Spécialité de doctorat : Ingénierie, Mécanique et Energétique

Thèse présentée et soutenue à Palaiseau, le 12 Décembre 2022, par

**MR. NAVEEN GOUTHAM**

Composition du Jury :

Dr. Albert Soret Group Leader, Barcelona Supercomputing Center	President & Reviewer
Dr. David Brayshaw Professor and Research Director, University of Reading	Reviewer
Dr. Aurélie Fischer Lecturer, Université de Paris	Examiner
Dr. Daniela Domeisen Assistant Professor, ETH Zurich	Examiner
Dr. Laurent Dubus Senior Scientist, RTE	Examiner
Dr. Lauriane Batté Research Manager, CNRM	Examiner
Dr. Philippe Drobinski Professor and Research Director, Ecole Polytechnique	Thesis supervisor
Dr. Hiba Omrani Expert Research Engineer, EDF	Thesis co-supervisor
Dr. Riwal Plougonven Professor, Ecole Polytechnique	Guest



There can be no creativity without constraints.  
— Unknown

To my mother. . .





# Acknowledgements

Foremost, I am grateful to Riwal Plougouven for offering me a PhD position, for adapting the subject based on my interests, for obtaining funding, and for guiding me to improve myself over the past years. I thank Philippe Drobinski for his co-supervision of my PhD, for his motivating words, and for his reactivity. I express my sincere gratitude to Bénédicte Jourdier and Laurent Dubus for initiating the relationship with EDF for my PhD. I thank Valérie Mériguet, my manager at EDF, for validating my PhD candidature at EDF, for her consistent support, and for all the opportunities that she has given me. I thank Hiba Omrani for her co-supervision of my PhD, and for her guidance, patience, and encouraging words throughout. I equally thank Sylvie Parey for co-supervising my PhD, for her energy, for helping me overcome my shyness in speaking French, and for being there whenever needed. I thank the ANRT (*Association Nationale de la Recherche et de la Technologie*) for the CIFRE (*Convention Industrielle de Formation par la Recherche*) fellowship for my PhD. I will be eternally grateful to my supervision team for trusting me with the subject despite my non-atmospheric sciences background and for adding fun to this journey.

I acknowledge the significant contributions of Alexis Tantet and Peter Tankov in the choice of mathematical methods. I thank Peter Hitchcock for initiating the internal database of sub-seasonal predictions at Laboratoire de Météorologie Dynamique. I acknowledge the data center ESPRI (*Ensemble de services pour la recherche à l'Institut Pierre-Simon Laplace*) for their help in storing and accessing the data. I equally acknowledge the people of the Energy4Climate Interdisciplinary Centre for organizing regular talks and creating a collaborative environment at LMD. I am grateful for the contributions of Albert Soret and Aurélie Fischer to the PhD committee. I thank Konstantinos M. Andreadis, Editor of the Journal of Hydrometeorology, for giving me an opportunity to review an article for the journal.

I thank my colleagues from the *Meteorology, Climate, and Renewable Energy Forecasting* group at EDF: Paul-Antoine, Stéphanie, Lila, Boutheina, Abdelhadi, Katy, Bruno, Benoit, Vincent, Lalaké, Shen, Romain, Christophe, Hajar, and Alexandra for their support. I acknowledge and appreciate the occasional chats with Jean-Baptiste. I thank my colleagues and friends at LMD: Assia and Clément for enriching my PhD life. I thank Patrick, Albert, and Cécile of LMD for their occasional encouraging talks. I express my gratitude to the IT support at both LMD (Julien and Karim) and EDF (Patricia, Karim, and Stéphane).

I thank all my interns: Boris, Ganglin, Omar, and Peng for their valuable contributions. Apart

## Chapter 0

---

from research, I also took part in other activities such as sewing and table tennis. I thank Lila for motivating me to learn sewing, and appreciate Augustin, Bruno, and Omar for challenging me in table tennis. I thank Camille of LMD for collaborations on the development of python packages to support open science. I express my gratitude to the people working in the canteens and cafeterias of Ecole Polytechnique and EDF for their warm welcome, and to the security and the cleaning staff of EDF for keeping the working place safe and clean.

I thank my friends: Amogh, Bageri, Vinay, Chiru, Pramod, Amar, Alija, Vania, Valentin, Amrita, and Mai for being there while navigating the difficult isolation times of COVID-19. I thank my always-backing family: dad Nagendran, sister Navya, and mother-in-law Leelavathi. Last but not the least, I express my sincere gratitude to my wife Keeru for supporting me throughout, for the good food, for motivating me when I was down, and for listening patiently to my blabber about the PhD. This PhD is dedicated to my mother Nalini who sacrificed everything for my education.

*Palaiseau, 12 December 2022*

Naveen Goutham.

# Abstract

Climate change has stimulated the energy sector, which is the largest emitter of global greenhouse gases (~40% in 2019), to transition to low-carbon energies. Europe, being one of the highest historical emitters of greenhouse gases, sits at the forefront of the energy transition. With a growing share of variable renewable power systems in the electricity mix on the one hand, and changing frequency and intensity of extreme events on the other, the weather-sensitive European energy sector is continuously on the lookout for accurate forecasts of essential climate variables on a continuum of timescales. The weather forecasts on short- to medium-range (i.e., from a few minutes ahead to at most two weeks) are reliable and essentially deterministic, and hence their operational use within the energy sector is well established. However, on timescales beyond two weeks and up to two months, i.e. in the sub-seasonal range, the predictions are necessarily probabilistic, and their reliability is far from that offered by short- and medium-range forecasts. Consequently, the operational use of sub-seasonal predictions within the energy sector is still in its infancy.

Having accurate information about the expected renewable energy production and electricity consumption on sub-seasonal timescales can help the energy sector in determining required reserve levels, scheduling maintenance, assessing and allocating risks attributed to extreme events, and estimating grid transmission capacity. In this regard, the main objective of this thesis is to provide more reliable information on sub-seasonal timescales, relative to climatology, to aid the energy sector in operational decision-making. We focus this research on 100-m wind speed and 2-m temperature over Europe.

As an essential first step, we rigorously assess the skill of sub-seasonal dynamical predictions of these two variables to quantify their predictability limits as they are delivered in a given forecasting system (the extended-range predictions of the European Centre for Medium-Range Weather Forecasts). We show that the weekly mean predictions of gridded temperature are more reliable than climatology for up to six weeks, and those of wind speed for up to three weeks. As a second step, we develop a statistical downscaling technique to reconstruct sub-seasonal predictions of wind speed and temperature using predictions of large-scale atmospheric circulation. We summarize the large-scale atmospheric state in a few indices by employing a dimension reduction methodology conditioned on wind speed and temperature over Europe. In other words, we use historical, observationally derived data to capture the relationship between the large-scale atmospheric circulation and our variables of interest (100 m wind speed and 2

m temperature). We then employ this relationship on sub-seasonal predictions of large-scale circulation, which are more reliable than surface variables, to deduce information about our variables of interest. This method allows us to produce, from a given ensemble of sub-seasonal predictions of large-scale circulation, a new ensemble of sub-seasonal predictions of our variables of interest. We demonstrate that the information thus extracted has value, as the hybrid ensemble combining both the dynamical and the statistical predictions of our variables of interest are more reliable than the dynamical predictions alone. As a final study, we investigate episodes of wind drought over Europe, because of their importance to the energy sector. A case study of the July 2018 episode of weak winds and the associated predictions, with and without our statistical downscaling methodology, illustrates the persistent difficulties of sub-seasonal predictions in predicting extreme events, in this case, due to the long-lasting challenge of forecasting blocking events.

Keywords: sub-seasonal predictions, wind speed, temperature, forecast verification, statistical predictions, Europe

# Résumé

Le changement climatique a incité le secteur de l'énergie, qui est le principal émetteur de gaz à effet de serre au niveau mondial (~40% en 2019), à se transitionner vers des énergies à faible teneur en carbone. Dans le cadre de la transition énergétique, la part des énergies renouvelables dans le mix énergétique est de plus en plus importante, rendant le système électrique plus sensible aux conditions météorologiques. En conséquence, le secteur de l'énergie est continuellement à la recherche de prévisions les plus précises possibles des variables climatiques sur un ensemble d'échelles de temps. Les prévisions météorologiques déterministes à court et moyen terme (de quelques minutes à deux semaines maximum) sont fiables, et leur utilisation opérationnelle dans le secteur de l'énergie est donc bien établie. Cependant, sur des échelles de temps infra-saisonniers, c'est à dire au-delà de deux semaines et jusqu'à deux mois, les prévisions sont nécessairement probabilistes, et leur fiabilité reste limitée. Par conséquent, l'utilisation opérationnelle des prévisions infra-saisonniers dans le secteur de l'énergie en est encore à ses débuts.

Disposer d'informations précises sur la production d'énergie renouvelable et la consommation d'électricité attendues sur des échelles de temps infra-saisonniers peut apporter une vraie valeur ajoutée au secteur de l'énergie. Les applications les plus importantes des prévisions infra-saisonniers sont la détermination des niveaux de réserve requis, la programmation de la maintenance, l'évaluation et la répartition des risques attribués aux événements extrêmes et l'estimation de la capacité de transmission du réseau. De ce fait, l'objectif principal de cette thèse est d'évaluer en premier temps et d'améliorer ensuite les prévisions infra-saisonniers par rapport à la climatologie, afin d'apporter des informations utiles et fiables au secteur de l'énergie. Nous nous concentrons dans ce travail sur la vitesse du vent à 100 m et la température à 2 m sur l'Europe.

Dans un premier temps, nous avons évalué les prévisions dynamiques infra-saisonniers en termes de vent et de température afin de quantifier leurs performances telles qu'elles sont fournies par le modèle de prévision. Nous avons montré que les prévisions de la température moyenne hebdomadaire sont plus fiables que la climatologie jusqu'à six semaines, et que celles de la vitesse du vent le sont jusqu'à trois semaines. Dans un deuxième temps, nous avons développé une technique de descente d'échelle statistique pour reconstruire des prévisions infra-saisonniers de la vitesse du vent et de la température en utilisant les prévisions de variables climatiques de grande échelle. Pour ce faire, nous avons utilisé des données historiques observées pour estimer la

relation entre la circulation atmosphérique à grande échelle et nos variables d'intérêt. Nous avons appliqué par la suite cette relation sur les prévisions infra-saisonniers de la circulation à grande échelle, qui sont plus fiables que celles des variables de surface, pour en déduire des prévisions de nos variables d'intérêt. Cette méthode nous a permis de produire, à partir des prévisions infra-saisonniers de la circulation à grande échelle, un nouvel ensemble de prévisions statistiques de température et de vent. Nous avons démontré que l'ensemble dit « hybride » combinant à la fois les nouvelles prévisions statistiques et les prévisions dynamiques de nos variables d'intérêt est plus fiable que les prévisions dynamiques seules. Pour la dernière partie de la thèse, nous avons développé une étude de cas sur les épisodes de faible vent en Europe, en raison de leur importance pour le secteur de l'énergie. Nous nous sommes intéressés à l'épisode de vents faibles de juillet 2018 et les prévisions associées. Pour cet événement, ni les prévisions dynamiques ni les prévisions statistiques n'ont réussi à le prévoir et ce en raison de la difficulté que les modèles de prévisions météorologiques ont à prévoir correctement les situations de blocage très souvent à l'origine de ces faibles vents.

Mots clés : prévisions infra-saisonniers, vitesse du vent, température, vérification des prévisions, prévisions statistiques, Europe

# Contents

<b>Acknowledgements</b>	<b>i</b>
<b>Abstract (English/French)</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Climate change and the energy sector . . . . .	1
1.2 Numerical Weather Prediction for the energy sector . . . . .	2
1.3 Addressing uncertainties in sub-seasonal predictions . . . . .	5
1.4 What constitutes a "good" sub-seasonal prediction? . . . . .	5
1.4.1 Metrics to assess the quality of sub-seasonal predictions . . . . .	6
1.4.2 The need for bias correction . . . . .	6
1.4.3 Improving sub-seasonal predictions of ECVs . . . . .	7
1.5 Objectives of the thesis . . . . .	7
1.6 Organization of the thesis manuscript . . . . .	8
<b>2 Quantitative assessment of sub-seasonal predictions</b>	<b>17</b>
<b>3 Improving sub-seasonal predictions</b>	<b>35</b>
<b>4 Case study: How well in advance can we anticipate wind droughts?</b>	<b>63</b>
<b>5 Conclusions and Perspectives</b>	<b>79</b>
5.1 Conclusions . . . . .	79
5.2 Perspectives . . . . .	80
5.2.1 Preliminary application . . . . .	81





# 1 Introduction

Modern Technology owes ecology an apology.  
— Alan M. Eddison, "Worse Verse" (1969)

## 1.1 Climate change and the energy sector

Anthropogenic climate change is one of the greatest threats modern humans have faced (Edenhofer et al., 2014). In 2019, more than 40% of the world's greenhouse gas emissions were attributed to heating and electricity production (IEA, 2021a). Nevertheless, to minimize the severe consequences of climate change, a collective global effort is in place to limit the global temperature rise to within 1.5° C with respect to pre-industrial levels. Europe, being one of the largest electricity consumers and greenhouse gas emitters in the world, sits at the forefront of economic decarbonization and energy transition (IEA, 2021b). The European Union, through the European Green Deal, has set a target to achieve climate neutrality (or net zero emissions) by 2050 (IEA, 2020). To achieve this goal, the European Commission has set an interim target to meet 40% of its final energy demand through renewable energy by 2030, which is almost double the share it met in 2019 (IEA, 2021b).

The increasing share of weather-sensitive, variable renewable power systems in the energy

mix raises concerns about the security of supply. With wind power becoming the largest renewable source of electricity in the European Union (IEA, 2020), several recent episodes of prolonged periods of low winds have increased concerns about the security of supply further (NewScientist, 2018; TheGuardian, 2018). One such case is the wind drought of July 2018, when certain regions in Europe experienced one of the lowest average wind speeds, relative to climatology. Figure 1.1 shows the average and anomaly of 100-m wind speed for the week between 16-22 July 2018 over and around the North Sea. This region has the highest density of installed wind power in Europe at the time of writing this manuscript (WindEurope, 2022) (Figure 1.2). As can be observed from Figure 1.1.a, the wind speed values are generally lower than the *cut-in speed*<sup>1</sup> of a typical wind turbine. Figure 1.1.b shows widespread, pronounced negative anomalies (down to -50%), such that the average over the domain shown is -26%. Since the power produced from a wind turbine is proportional to the cube of wind speed, a 26% decrease in wind speed leads to about 60% reduction in wind power, putting the energy sector which is reliant on wind power at serious risk of undersupply. Since fossil fuel-based power systems are called upon to make up for the demand unfulfilled by

<sup>1</sup>The wind speed at which a wind turbine starts to generate power. The cut-in speed for a typical wind turbine is between 3 and 4 m s<sup>-1</sup> (Manwell et al., 2010; Burton et al., 2011).

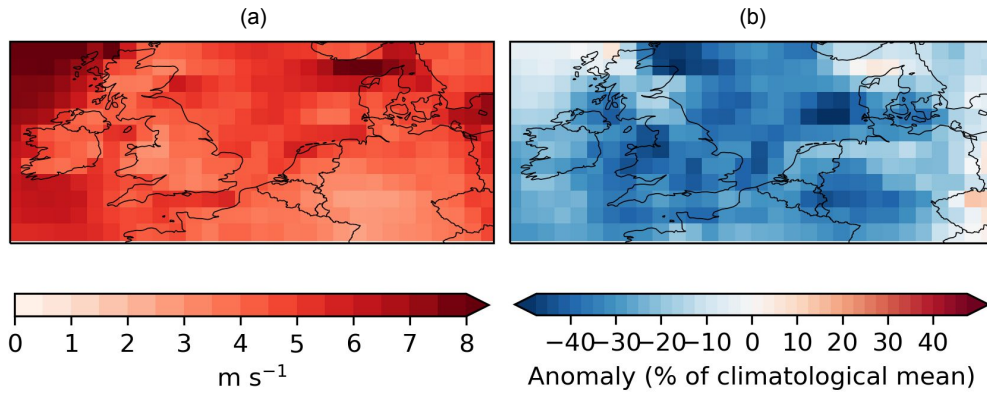


Figure 1.1: (a) Average 100-m wind speed over the North Sea for the week between 16-22 July 2018. (b) 100-m wind speed anomaly for the same week as a percentage of the climatological mean. The climatology corresponds to the mean wind speed in July between 1979 and 2021 leaving out 2018. The wind speed values are derived from ERA5 reanalysis.

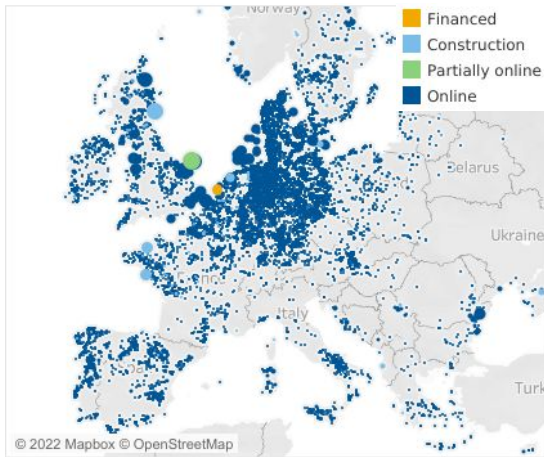


Figure 1.2: Location of the major onshore and offshore wind farms in Europe as of August 2022. The map is sourced from Wind Europe (<https://windeurope.org/>).

renewable systems, the energy sector requires accurate anticipation of the expected renewable production to sustain and support the future growth of renewables. Thus, the most important question that arises from such extreme cases is: How well in advance can the energy industry foresee such events to plan alternative solutions?

## 1.2 Numerical Weather Prediction for the energy sector

Meteorological predictions are indispensable for the smooth operation of the energy sector. As demand and supply are to be balanced at every instant, the energy sector with a high share of renewable systems requires accurate predictions of essential climate variables (ECVs) such as temperature, wind speed, solar radiation, and precipitation on a range of time scales. Numerical Weather Prediction (NWP), a method to predict the weather at a future period in time knowing the initial conditions and the laws that govern the evolution of weather (Bauer et al., 2015), is at the heart of renewable energy production and electricity demand forecasting. Although today's NWP models are more sophisticated in terms of representation of various components of the earth system and their corresponding interactions, and more advanced in terms of accurately predicting the weather than they used to be two decades ago, we still do not completely understand the complex interactions between various components of the Earth system. In addition, the initial conditions for

the NWP models, which are assimilated from several ground-, upper air-, or space-based observations, are only approximations. Hence, in practice, numerical weather predictions are carried out from imperfect initial conditions, using imperfect numerical formulations (Lorenz, 1963, 1982; Buizza et al., 2015). The convolution of amplified errors from initial conditions due to atmospheric instabilities with errors from the use of imperfect numerical models is what sets the forecasting *skill horizon*<sup>I</sup> (e.g., Robertson and Vitart, 2018).

The complexity of an NWP model depends on the forecasting lead time. The predictions are categorized based on the time horizon for which they are made. They are:

1. *Nowcasting* refers to the detailed description of current weather, and forecasting up to several hours ahead (e.g., WMO, 2017). Nowcasting is an initial condition problem that requires high-quality local observations and an NWP model to forecast the evolution of weather;
2. *Short-range forecasting* refers to forecasting weather from a few hours ahead and up to a few days (e.g., Doswell, 1986);
3. *Medium-range forecasting* involves predicting weather from a few days ahead and up to two weeks (e.g., Wagner, 1989). Similar to nowcasting, short-range and medium-range forecasting are also initial condition problems where the predictability is driven by atmospheric initial conditions. A majority of the NWP models used to produce nowcasts, short- and medium-range forecasts generally include only a

description of land and atmospheric processes (e.g., Bougeault et al., 2010);

4. *Seasonal predictions* involve predicting anomalies relative to climate normal on timescales ranging from a month ahead to several seasons. The predictability on seasonal timescales mainly stems from the variability of sea surface temperatures, especially the state of El-Nino Southern Oscillation (e.g., Kushnir et al., 2006; Doblas-Reyes et al., 2013). In addition, sea ice extent (e.g., Guemas et al., 2016) and land conditions (e.g., Prodhomme et al., 2016) also contribute to the predictability on seasonal timescales. As a result, the NWP models which produce seasonal predictions generally use atmosphere/land/ocean coupled models, with some of them using an interactive sea-ice model in addition (Robertson and Vitart, 2018).
5. *Decadal predictions* involve predictions on multi-annual to decadal timescales. The predictability on these timescales arises mainly from multi-annual climatic variability such as the Atlantic Multidecadal Oscillation (e.g., Dijkstra et al., 2006) and the Pacific Decadal Oscillation (e.g., Mantua and Hare, 2002). Furthermore, the predictability also originates from external forcings such as anthropogenic greenhouse gas emissions, aerosols, and volcanic emissions (e.g., Van Oldenborgh et al., 2012).

From the description of the categories of predictions, it is evident that there is a gap between medium-range forecasts (i.e., up to two weeks) and seasonal predictions (i.e., month ahead to several seasons). This timescale, i.e. from two weeks ahead to two months ahead, is referred to as the *sub-seasonal* timescale, and the predictability on this timescale comes from the

<sup>I</sup>The lead time after which the forecast error becomes larger than that of a baseline (e.g., climatology or persistence).

accurate description of both the atmospheric initial conditions and boundary forcings (Hoskins, 2012; Robertson and Vitart, 2018).

### Sub-seasonal predictions for the energy sector

Predictability on sub-seasonal timescales emanates predominantly from the Madden-Julian Oscillation (e.g., Jones et al., 2004b,a; Zheng et al., 2018), land conditions (e.g., Koster et al., 2011; van den Hurk et al., 2012; Prodhomme et al., 2016; Seo et al., 2019), ocean conditions (e.g., Woolnough et al., 2007; Fu et al., 2007; Subramanian et al., 2019), snow cover (e.g., Sobolowski et al., 2010; Lin and Wu, 2011; Orsolini et al., 2013), and stratosphere-troposphere interactions (e.g., Baldwin et al., 2003; Domeisen et al., 2020; Schwartz and Garfinkel, 2020). Because of the chaos inherent in the atmosphere, predictability of fine-scale spatio-temporal features is poor on sub-seasonal timescales (e.g., Lorenz, 1965; Jifan, 1989; Zhang et al., 2019). However, the predictability of large-scale, low-frequency ocean, land, and cryospheric features lasts well beyond two weeks (Vitart et al., 2012; Buizza and Leutbecher, 2015; Toth and Buizza, 2019). Hence, predictions on subseasonal timescales are generally averaged over a large enough spatiotemporal scale (e.g., weekly mean temperature over Belgium) to extract relevant and predictable components of the signal (Lorenz, 1982; Zhu et al., 2014; Buizza and Leutbecher, 2015).

Sub-seasonal timescale is instrumental for the energy industry in determining required reserve levels, scheduling maintenance, assessing risks from extreme events and planning mitigation strategies, estimating grid transmission capacities, and hedging against risks in the derivative markets. Currently, the energy industry mainly

uses information based on climatology to make decisions on sub-seasonal timescales. With increasing renewable power in the electricity mix, on the one hand, (IEA, 2020), and changing frequency and intensity of extreme events on the other (Seneviratne et al., 2012), there is a growing demand for sub-seasonal predictions of renewable production and electricity consumption from the energy sector for operational decision-making.

The energy sector relies on standard meteorological variables to estimate renewable energy production and electricity consumption. The key variable used to estimate electricity consumption is the surface temperature. Even though there exist several peer-reviewed research on the verification of sub-seasonal predictions of 2 m temperature, they are either limited to some ground-based stations across Europe (e.g., Monhart et al., 2018), or a specific geographic domain (e.g., Vigaud et al., 2019), or country-wide averages (e.g., Dorrington et al., 2020). The fundamental variable used to estimate wind energy production is wind speed. Until recently, the energy sector was accustomed to wind speed at 10 m extrapolated vertically to the turbine hub height to estimate wind energy production (Manwell et al., 2010; Burton et al., 2011). Since the vertical extrapolation of wind speed from 10 m to the turbine hub height could lead to significant errors (e.g. Jourdier, 2015), the weather centers around the world started archiving wind speed at 100 m, based on the demand of the energy sector, as it is close to the hub height of a typical wind turbine (WindEurope, 2022). Although several research works have looked at sub-seasonal predictions of 10 m wind speed in the past (e.g., Lynch et al., 2014; Büeler et al., 2020; Lledó and Doblás-Reyes, 2020), the author, based on his research, found no published peer-reviewed research on the verification of

sub-seasonal predictions of 100 m wind speed at the time of commencement of this Ph.D. (i.e., January 2020). Hence, to fill the gap in research, this Ph.D. thesis will focus on gridded sub-seasonal predictions of 2-m temperature and 100-m wind speed over Europe.

### 1.3 Addressing uncertainties in sub-seasonal predictions

Every prediction made using NWP models suffers from two kinds of errors: initial condition errors and model errors. As the earth-based observations are spatially discontinuous, weather centers around the world assimilate data they receive from several ground-, upper air-, or space-based observations to a previously run short-range forecast to create a 3-dimensional spatially-homogeneous estimate of the atmospheric state. This estimate of the atmospheric state is known as *analysis*. All the NWP models are initialized from the analysis. Since the atmosphere is chaotic, even small errors in initial conditions could lead to significant errors in forecasts.

The most important model related errors in numerical weather prediction are due to missing physical processes and parametrizations (e.g., Palmer et al., 2009; Julia and Tim, 2011; Palmer, 2012; Leutbecher et al., 2016; Robertson and Vitart, 2018; McTaggart-Cowan et al., 2022). An additional source of model related errors is from the use of limited spatial resolution in NWP models. The complex convolution of errors from initial conditions with model errors leads to uncertainties in predictions, which amplify with increasing lead time.

The initial condition and model uncertainties are taken into account in the NWP models by producing an ensemble of numerical integrations

instead of just one best estimate of the future state (Buizza, 2019). Then, the future state is a range of possibilities with varying probabilities. This shift from determinism to probabilism has been one of the greatest breakthroughs in numerical weather prediction (Palmer, 2012). The uncertainties in initial conditions in NWP models are addressed by perturbing the best estimated initial state using ensemble and/or variational methods (refer to Bannister (2017) for a review of all the available methods). The model uncertainty is taken into account through the use of multi-model, perturbed parameter, perturbed tendency, or stochastic backscatter approaches (e.g., Buizza, 2019; McTaggart-Cowan et al., 2022). Comparison of different methods that take into account uncertainties in initial conditions and model formulations is important for advancing model development. Nevertheless, this is out of the scope of this thesis.

### 1.4 What constitutes a "good" sub-seasonal prediction?

How can the energy sector know whether the information obtained from sub-seasonal predictions is more reliable than what is already available at their fingertips (i.e., climatology)? As the information on sub-seasonal timescales is available in the form of probabilities, the energy sector, which is conventionally used to a deterministic way of treating forecasts on short- to medium-range timescales, finds it challenging to extract pertinent information from an ensemble prediction. Irrespective of the forecasting time horizon, predictions are evaluated by assessing different quality attributes (Jolliffe and Stephenson, 2003; Coelho et al., 2019; Wilks, 2019). The most important prediction quality attributes are:

- *Accuracy*: a measure of the average distance between forecasts and observations;
- *Association*: a measure of the strength of the relationship between forecasts and observations;
- *Reliability*: a measure of calibration of the issued forecast probabilities;
- *Resolution*: a measure of changes in the frequency of occurrence of an event as a function of variation in the issued forecast probabilities;
- *Discrimination*: a measure of the ability of the forecasting system to produce forecasts that discriminate between events and non-events. It is closely related to resolution;
- *Sharpness*: a measure of the ability of forecasts to produce concentrated predictive distributions which are distinct from climatological probabilities.

#### 1.4.1 Metrics to assess the quality of sub-seasonal predictions

Several scores have been proposed in the literature to measure the quality of sub-seasonal ensemble predictions taking into account different forecast attributes (Jolliffe and Stephenson, 2003; Coelho et al., 2019; Wilks, 2019). The most widely used scores in forecast verification are the Anomaly Correlation Coefficient (Namias, 1952; Wilks, 2019), the Brier Score (Brier, 1950; Murphy, 1973), the Ranked Probability Score (Murphy, 1970), and the Continuous Ranked Probability Score (Matheson and Winkler, 1976; Unger, 1985; Hersbach, 2000). The anomaly Correlation Coefficient measures the strength of the relationship (i.e., association) between observed and forecast ensemble mean anomalies by ignoring systematic biases.

While Brier Score measures the accuracy of forecasts for a dichotomous event, the Ranked Probability Score measures the accuracy for a multi-event situation. The Continuous Ranked Probability Score is an extension of the Ranked Probability Score where it measures the accuracy of a forecast for an infinite-event situation of infinitesimal width (Wilks, 2019). All three accuracy metrics are proper scores which can be further decomposed into components representing reliability and resolution (Murphy, 1970, 1973; Hersbach, 2000).

Apart from pure mathematical scores, a lot of progress has been made toward the development of graphical products to help visualize various forecast attributes. Notable among these are Reliability diagram (Sanders, 1963; Wilks, 2019) and Relative Operating Characteristic diagram (Mason, 1982; Harvey et al., 1992). A reliability diagram is used to visualize the reliability, resolution, and sharpness attributes of forecasts for dichotomous events (i.e., yes/no events). A Relative Operating Characteristic diagram is used to visualize the discrimination attribute of the forecasting system. In this thesis, we will use several metrics and graphical tools to understand the behavior of different attributes of sub-seasonal predictions to obtain a comprehensive overview of prediction quality.

#### 1.4.2 The need for bias correction

Sub-seasonal predictions, even upon incorporating methods to account for initial condition and model-related uncertainties in the NWP models which produce them, suffer from random and systematic errors. While random errors are unpredictable, systematic errors can be anticipated to a certain extent (Siegert and Stephenson, 2019).

Systematic errors are generally in the form of consistent over- or under-estimation of values, over- or under-dispersion of the distribution, or changes in the properties of other moments of the distribution relative to observations (Siegert and Stephenson, 2019). Systematic errors in sub-seasonal predictions can be corrected through statistical post-processing techniques such as mean and variance adjustment, quantile mapping, or calibration (e.g., Jolliffe and Stephenson, 2003; Gneiting et al., 2007; Siegert and Stephenson, 2019; Manzanas et al., 2019). Each method has its advantages and disadvantages, and the choice of the method depends on the end-user needs.

All the methods, however, require sufficiently long, historical, paired prediction-observation data to learn and correct the errors. The weather centers around the world which produce operational sub-seasonal predictions also produce a suite of historical sub-seasonal predictions, generated using the same operational model for a period in the past, known as retrospective forecasts (or re-forecasts) (Buizza, 2019). These re-forecasts serve two purposes: one, to estimate the climatology of the operational sub-seasonal prediction model; two, to correct systematic errors of operational predictions. Furthermore, re-forecasts can be used to study the long-term performance of the model. It is imperative to correct systematic errors of sub-seasonal predictions before employing them for research or other practical applications.

### 1.4.3 Improving sub-seasonal predictions of ECVs

Despite applying bias-correction techniques to correct for systematic errors prevalent in sub-seasonal predictions, the predictions of surface fields (i.e., within the planetary boundary layer)

are hardly better than a baseline climatology (e.g. Lynch et al., 2014; Monhart et al., 2018; Vigaud et al., 2019). This, in addition to chaos inherent in the atmosphere, is due to the complex convolution of initial condition and model errors with increasing lead time, which result in the misrepresentation of physical relationships between large-scale, low-frequency fields (e.g., upper-level geopotential height, sea surface temperature, etc.) and surface fields (Palmer et al., 2009; Leutbecher et al., 2016; Robertson and Vitart, 2018; Lledó and Doblas-Reyes, 2020). As the surface fields are more sensitive to model parametrizations compared to large-scale fields, the skill horizon of large-scale fields is longer than that of surface fields (Buizza and Leutbecher, 2015; Toth and Buizza, 2019). Given the physical relationships between large-scale and surface fields, and the longer skill horizon of large-scale fields compared to surface fields, we can apply statistical downscaling methods on dynamical predictions of large-scale fields to derive statistical sub-seasonal predictions of surface fields (Alonzo et al., 2017; Alonzo, 2018; Manzanas et al., 2018; Grams et al., 2017; Goutham et al., 2021; Ramon et al., 2021). The work of Alonzo et al. (2017) on statistical downscaling of seasonal predictions, carried out previously at Laboratoire de Météorologie Dynamique, has shown a promising potential. However, their implementation did not allow for a comparison with the dynamical seasonal predictions (i.e., off-the-shelf predictions) of surface fields. Hence, the work plan of this thesis takes a different approach as described in the section 1.6.

## 1.5 Objectives of the thesis

This thesis aims to provide reliable information about the expected temperature and wind speed



on sub-seasonal timescales over Europe for the energy sector. In this regard, the thesis is guided by the following key questions:

1. Are the available (i.e., off-the-shelf) European sub-seasonal predictions of 100-m wind speed and 2-m temperature more reliable than baseline climatology?
2. How to improve the European sub-seasonal predictions of the same two variables?
3. How well in advance can we anticipate extreme events with significant impact on the energy sector?

## 1.6 Organization of the thesis manuscript

This thesis is presented as a compilation of three articles. **Chapter 2** tackles the first objective where we carry out a comprehensive quality assessment of off-the-shelf, gridded sub-seasonal predictions of 100-m wind speed and 2-m temperature over Europe using several forecast quality metrics. This work has been published in the *Monthly Weather Review* in June 2022.

In **chapter 3**, we develop a novel statistical downscaling methodology to improve the sub-seasonal predictions of 100-m wind speed and 2-m temperature over Europe. This work has been accepted for publication in the *Monthly Weather Review* in October 2022.

**Chapter 4** focuses on extreme events of significant importance to the energy sector such as wind droughts to determine the time horizon before which we could anticipate such events. The manuscript of this chapter has been submitted to *Monthly Weather Review* in October 2022, and is currently under review. The main

conclusions and future perspectives of the thesis are presented in **chapter 5**.

## References

- Alonzo, B., 2018: Seasonal forecasting of wind energy resource and production in France and associated risk. Ph.D. thesis, Université Paris-Saclay.
- Alonzo, B., H.-K. Ringkjøb, B. Jourdier, P. Drobinski, R. Plougonven, and P. Tankov, 2017: Modelling the variability of the wind energy resource on monthly and seasonal timescales. *Renewable energy*, **113**, 1434–1446.
- Baldwin, M. P., D. B. Stephenson, D. W. J. Thompson, T. J. Dunkerton, A. J. Charlton, and A. O’Neill, 2003: Stratospheric Memory and Skill of Extended-Range Weather Forecasts. *Science*, **301 (5633)**, 636–640, <https://doi.org/10.1126/science.1087143>, publisher: American Association for the Advancement of Science Section: Report.
- Bannister, R. N., 2017: A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **143 (703)**, 607–633, <https://doi.org/10.1002/qj.2982>.
- Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, <https://doi.org/10.1038/nature14956>.
- Bougeault, P., Z. Toth, C. Bishop, B. Brown, D. Burridge, D. H. Chen, B. Ebert, M. Fuentes, T. M. Hamill, K. Mylne, J. Nicoulaou, T. Paccagnella, Y.-Y. Park, D. Parsons,

- B. Raoult, D. Schuster, P. S. Dias, R. Swinbank, Y. Takeuchi, W. Tennant, L. Wilson, and S. Worley, 2010: The THORPEX interactive grand global ensemble. *Bulletin of the American Meteorological Society*, **91** (8), 1059 – 1072, <https://doi.org/10.1175/2010BAMS2853.1>.
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, **78** (1), 1 – 3, [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2).
- Buizza, R., 2019: Ensemble generation: The TIGGE and S2S ensembles. *Sub-Seasonal to Seasonal Prediction: The Gap between Weather and Climate Forecasting*, A. Robertson, and F. Vitart, Eds., Elsevier, chap. 13, 261–303.
- Buizza, R., and M. Leutbecher, 2015: The forecast skill horizon. *Quarterly Journal of the Royal Meteorological Society*, **141** (693), 3366–3382, <https://doi.org/10.1002/qj.2619>.
- Buizza, R., M. Leutbecher, and A. Thorpe, 2015: Living with the butterfly effect: A seamless view of predictability. *ECMWF newsletter*, **145**, 18–23.
- Burton, T., N. Jenkins, D. Sharpe, and E. Bossanyi, 2011: *Wind energy handbook*. John Wiley & Sons.
- Büeler, D., R. Beerli, H. Wernli, and C. M. Grams, 2020: Stratospheric influence on ECMWF sub-seasonal forecast skill for energy-industry-relevant surface weather in European countries. *Quarterly Journal of the Royal Meteorological Society*, **146** (733), 3675–3694, <https://doi.org/10.1002/qj.3866>.
- Coelho, C. A. S., B. Brown, L. Wilson, M. Mittermaier, and B. Casati, 2019: Forecast verification for S2S timescales. *Sub-Seasonal to Seasonal Prediction: The Gap between Weather and Climate Forecasting*, A. Robertson, and F. Vitart, Eds., Elsevier, 337–361, section: 16.
- Dijkstra, H., L. Te Raa, M. Schmeits, and J. Gerrits, 2006: On the physics of the Atlantic Multidecadal Oscillation. *Ocean Dynamics*, **56**, 36–50, <https://doi.org/10.1007/s10236-005-0043-0>.
- Doblas-Reyes, F. J., J. García-Serrano, F. Lienert, A. P. Biescas, and L. R. L. Rodrigues, 2013: Seasonal climate predictability and forecasting: status and prospects. *WIREs Climate Change*, **4** (4), 245–268, <https://doi.org/10.1002/wcc.217>.
- Domeisen, D. I. V., A. H. Butler, A. J. Charlton-Perez, B. Ayarzagüena, M. P. Baldwin, E. Dunn-Sigouin, J. C. Furtado, C. I. Garfinkel, P. Hitchcock, A. Y. Karpechko, H. Kim, J. Knight, A. L. Lang, E.-P. Lim, A. Marshall, G. Roff, C. Schwartz, I. R. Simpson, S.-W. Son, and M. Taguchi, 2020: The role of the stratosphere in subseasonal to seasonal prediction: 2. predictability arising from stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmospheres*, **125** (2), e2019JD030923, <https://doi.org/10.1029/2019JD030923>.
- Dorrington, J., I. Finney, T. Palmer, and A. Weisheimer, 2020: Beyond skill scores: exploring sub-seasonal forecast value through a case-study of French month-ahead energy prediction. *Quarterly Journal of the Royal Meteorological Society*, **146** (733), 3623–3637, <https://doi.org/10.1002/qj.3863>.
- Doswell, C. A., 1986: *Short-Range Forecasting*, 689–719. American Meteorological Society, Boston, MA, [https://doi.org/10.1007/978-1-935704-20-1\\_29](https://doi.org/10.1007/978-1-935704-20-1_29).

- Edenhofer, O., R. Pichs-Madruga, Y. Sokona, E. Farahani, K. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel, and J. M. (eds.), 2014: Climate change 2014: Mitigation of climate change. contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.
- Fu, X., B. Wang, D. E. Waliser, and L. Tao, 2007: Impact of Atmosphere–Ocean Coupling on the Predictability of Monsoon Intraseasonal Oscillations. *Journal of the Atmospheric Sciences*, **64** (1), 157–174, <https://doi.org/10.1175/JAS3830.1>, publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences.
- Gneiting, T., F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69** (2), 243–268, <https://doi.org/10.1111/j.1467-9868.2007.00587.x>.
- Goutham, N., B. Alonzo, A. Dupré, R. Plougonven, R. Doctors, L. Liao, M. Mougeot, A. Fischer, and P. Drobinski, 2021: Using machine-learning methods to improve surface wind speed from the outputs of a numerical weather prediction model. *Boundary-Layer Meteorology*, **179** (1), 133–161.
- Grams, C. M., R. Beerli, S. Pfenninger, I. Staffell, and H. Wernli, 2017: Balancing Europe’s wind-power output through spatial deployment informed by weather regimes. *Nature Climate Change*, **7** (8), 557–562, <https://doi.org/10.1038/nclimate3338>, number: 8 Publisher: Nature Publishing Group.
- Guemas, V., M. Chevallier, M. Déqué, O. Bellprat, and F. Doblas-Reyes, 2016: Impact of sea ice initialization on sea ice and atmosphere prediction skill on seasonal timescales. *Geophysical Research Letters*, **43** (8), 3889–3896, <https://doi.org/10.1002/2015GL066626>.
- Harvey, L. O., K. R. Hammond, C. M. Lusk, and E. F. Mross, 1992: The application of signal detection theory to weather forecasting behavior. *Monthly Weather Review*, **120** (5), 863 – 883, [https://doi.org/10.1175/1520-0493\(1992\)120<0863:TAOSDT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<0863:TAOSDT>2.0.CO;2).
- Hersbach, H., 2000: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting*, **15** (5), 559 – 570, [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2), place: Boston MA, USA Publisher: American Meteorological Society.
- Hoskins, B., 2012: Predictability Beyond the Deterministic Limit. World Meteorological Organization, URL <https://public.wmo.int/en/bulletin/predictability-beyond-deterministic-limit>.
- IEA, 2020: European Union 2020. Tech. rep., Paris, France. URL <https://www.iea.org/reports/european-union-2020>.
- IEA, 2021a: Greenhouse gas emissions from energy: Overview. Tech. rep., Paris, France. URL <https://tinyurl.com/2pt8ea4a>.
- IEA, 2021b: Net Zero by 2050. Tech. rep., Paris, France. URL <https://www.iea.org/reports/net-zero-by-2050>.
- Jifan, C., 1989: Predictability of the atmosphere. *Advances in Atmospheric Sciences*, **6** (3), 335–346, <https://doi.org/10.1007/BF02661539>.
- Jolliffe, I., and D. Stephenson, 2003: *Forecast Verification: A Practitioner’s Guide in Atmo-*

- spheric Science*. Wiley, URL <https://books.google.com/books?id=cjS9kK8IWbwC>.
- Jones, C., D. E. Waliser, K. M. Lau, and W. Stern, 2004a: Global Occurrences of Extreme Precipitation and the Madden–Julian Oscillation: Observations and Predictability. *Journal of Climate*, **17** (23), 4575–4589, <https://doi.org/10.1175/3238.1>, publisher: American Meteorological Society Section: Journal of Climate.
- Jones, C., D. E. Waliser, K. M. Lau, and W. Stern, 2004b: The Madden–Julian Oscillation and Its Impact on Northern Hemisphere Weather Predictability. *Monthly Weather Review*, **132** (6), 1462–1471, [https://doi.org/10.1175/1520-0493\(2004\)132<1462:TMOAII>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1462:TMOAII>2.0.CO;2), publisher: American Meteorological Society Section: Monthly Weather Review.
- Jourdier, B., 2015: Wind resource in metropolitan France: assessment methods, variability and trends. Ph.D. thesis, Ecole Polytechnique, Palaiseau, France, URL [http://inis.iaea.org/search/search.aspx?orig\\_q=RN:48072202](http://inis.iaea.org/search/search.aspx?orig_q=RN:48072202).
- Julia, S., and P. Tim, 2011: Uncertainty in weather and climate prediction. *Phil. Trans. R. Soc. A*, **369**, 4751–4767, <https://doi.org/10.1098/rsta.2011.0161>.
- Koster, R. D., S. P. P. Mahanama, T. J. Yamada, G. Balsamo, A. A. Berg, M. Boisserie, P. A. Dirmeyer, F. J. Doblas-Reyes, G. Drewitt, C. T. Gordon, Z. Guo, J.-H. Jeong, W.-S. Lee, Z. Li, L. Luo, S. Malyshev, W. J. Merryfield, S. I. Seneviratne, T. Stanelle, B. J. J. M. v. d. Hurk, F. Vitart, and E. F. Wood, 2011: The Second Phase of the Global Land–Atmosphere Coupling Experiment: Soil Moisture Contributions to Sub-seasonal Forecast Skill. *Journal of Hydrometeorology*, **12** (5), 805–822, <https://doi.org/10.1175/2011JHM1365.1>, publisher: American Meteorological Society Section: Journal of Hydrometeorology.
- Kushnir, Y., W. A. Robinson, P. Chang, and A. W. Robertson, 2006: The physical basis for predicting Atlantic sector seasonal-to-interannual climate variability. *Journal of Climate*, **19** (23), 5949–5970, <https://doi.org/10.1175/JCLI3943.1>.
- Leutbecher, M., P. Ollinaho, S.-J. Lock, S. Lang, P. Bechtold, A. Beljaars, A. Bozzo, R. Forbes, T. Haiden, R. Hogan, and I. Sandu, 2016: Model uncertainty representations in the IFS. *ECMWF/WWRP Workshop: Model Uncertainty*, ECMWF, Reading, URL <https://www.ecmwf.int/node/16369>.
- Lin, H., and Z. Wu, 2011: Contribution of the Autumn Tibetan Plateau Snow Cover to Seasonal Prediction of North American Winter Temperature. *Journal of Climate*, **24** (11), 2801–2813, <https://doi.org/10.1175/2010JCLI3889.1>, publisher: American Meteorological Society Section: Journal of Climate.
- Lledó, L., and F. J. Doblas-Reyes, 2020: Predicting Daily Mean Wind Speed in Europe Weeks ahead from MJO Status. *Monthly Weather Review*, **148** (8), 3413–3426, <https://doi.org/10.1175/MWR-D-19-0328.1>.
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *Journal of atmospheric sciences*, **20** (2), 130–141.
- Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17** (3), 321–333, <https://doi.org/10.1111/j.2153-3490.1965.tb01424.x>.

- Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34** (6), 505–513.
- Lynch, K. J., D. J. Brayshaw, and A. Charlton-Perez, 2014: Verification of European Sub-seasonal Wind Speed Forecasts. *Monthly Weather Review*, **142**, 13.
- Mantua, N., and S. Hare, 2002: The Pacific Decadal Oscillation. *Journal of Oceanography*, **58**, 35–44, <https://doi.org/10.1023/A:1015820616384>.
- Manwell, J. F., J. G. McGowan, and A. L. Rogers, 2010: *Wind energy explained: theory, design and application*. John Wiley & Sons.
- Manzanas, R., J. Gutiérrez, J. Fernández, E. van Meijgaard, S. Calmanti, M. Magariño, A. Cofiño, and S. Herrera, 2018: Dynamical and statistical downscaling of seasonal temperature forecasts in Europe: Added value for user applications. *Climate Services*, **9**, 44–56, <https://doi.org/10.1016/j.cliser.2017.06.004>, climate services in practice: what we learnt from EUPORIAS.
- Manzanas, R., J. M. Gutiérrez, J. Bhend, S. Hemri, F. J. Doblas-Reyes, V. Torralba, E. Penabad, and A. Brookshaw, 2019: Bias adjustment and ensemble recalibration methods for seasonal forecasting: a comprehensive intercomparison using the C3S dataset. *Climate Dynamics*, **53** (3-4), 1287–1305, <https://doi.org/10.1007/s00382-019-04640-4>.
- Mason, I., 1982: A model for assessment of weather forecasts. *Australian Meteorological Magazine*, **30**, 291–303, URL <http://www.bom.gov.au/jshess/docs/1982/mason.pdf>.
- Matheson, J. E., and R. L. Winkler, 1976: Scoring Rules for Continuous Probability Distributions. *Management Science*, **22** (10), 1087–1096, URL [https://econpapers.repec.org/article/inmormnsc/v\\_3a22\\_3ay\\_3a1976\\_3ai\\_3a10\\_3ap\\_3a1087-1096.htm](https://econpapers.repec.org/article/inmormnsc/v_3a22_3ay_3a1976_3ai_3a10_3ap_3a1087-1096.htm), publisher: INFORMS.
- McTaggart-Cowan, R., L. Separovic, R. Aider, M. Charron, M. Desgagné, P. L. Houtekamer, D. Paquin-Ricard, P. A. Vaillancourt, and A. Zadra, 2022: Using Stochastically Perturbed Parameterizations to represent model uncertainty, part I: Implementation and parameter sensitivity. *Monthly Weather Review*, <https://doi.org/10.1175/MWR-D-21-0315.1>.
- Monhart, S., C. Spirig, J. Bhend, K. Bogner, C. Schär, and M. A. Liniger, 2018: Skill of Subseasonal Forecasts in Europe: Effect of Bias Correction and Downscaling Using Surface Observations. *Journal of Geophysical Research: Atmospheres*, <https://doi.org/10.1029/2017JD027923>.
- Murphy, A. H., 1970: The Ranked Probability Score and The Probability Score: A Comparison. *Monthly Weather Review*, **98** (12), 917 – 924, [https://doi.org/10.1175/1520-0493\(1970\)098<0917:TRPSAT>2.3.CO;2](https://doi.org/10.1175/1520-0493(1970)098<0917:TRPSAT>2.3.CO;2).
- Murphy, A. H., 1973: A new vector partition of the probability score. *Journal of Applied Meteorology and Climatology*, **12** (4), 595 – 600, [https://doi.org/10.1175/1520-0450\(1973\)012<0595:ANVPOT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2).
- Namias, J., 1952: The Annual Course of Month-to-Month Persistence in Climatic Anomalies. *Bulletin of the American Meteorological Society*, **33** (7), 279–285, <https://doi.org/10.1175/1520-0477-33.7.279>.

- NewScientist, 2018: Weird 'wind drought' means Britain's turbines are at a standstill. URL <https://tinyurl.com/4euexhv3>.
- Orsolini, Y. J., R. Senan, G. Balsamo, F. J. Doblas-Reyes, F. Vitart, A. Weisheimer, A. Carrasco, and R. E. Benestad, 2013: Impact of snow initialization on sub-seasonal forecasts. *Climate Dynamics*, **41** (7), 1969–1982, <https://doi.org/10.1007/s00382-013-1782-0>.
- Palmer, T. N., 2012: Towards the probabilistic Earth-system simulator: a vision for the future of climate and weather prediction. *Quarterly Journal of the Royal Meteorological Society*, **138** (665), 841–861, <https://doi.org/10.1002/qj.1923>.
- Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. (598), 42, URL <https://www.ecmwf.int/node/11577>, publisher: ECMWF.
- Prodhomme, C., F. Doblas-Reyes, O. Bellprat, and E. Dutra, 2016: Impact of land-surface initialization on sub-seasonal to seasonal forecasts over Europe. *Climate Dynamics*, **47**, 919–935, <https://doi.org/10.1007/s00382-015-2879-4>.
- Ramon, J., L. Lledó, P.-A. Bretonnière, M. Samsó, and F. J. Doblas-Reyes, 2021: A perfect prognosis downscaling methodology for seasonal prediction of local-scale wind speeds. *Environmental Research Letters*, **16** (5), 054010, <https://doi.org/10.1088/1748-9326/abe491>.
- Robertson, A., and F. Vitart, 2018: *Sub-seasonal to Seasonal Prediction: The Gap Between Weather and Climate Forecasting*. Elsevier.
- Sanders, F., 1963: On Subjective Probability Forecasting. *Journal of Applied Meteorology and Climatology*, **2** (2), 191–201, [https://doi.org/10.1175/1520-0450\(1963\)002<0191:OSPF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1963)002<0191:OSPF>2.0.CO;2), publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.
- Schwartz, C., and C. I. Garfinkel, 2020: Troposphere-stratosphere coupling in subseasonal-to-seasonal models and its importance for a realistic extratropical response to the Madden-Julian Oscillation. *Journal of Geophysical Research: Atmospheres*, **125** (10), e2019JD032043, <https://doi.org/10.1029/2019JD032043>.
- Seneviratne, S., N. Nicholls, D. Easterling, C. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, M. Rahimi, M. Reichstein, A. Sorteberg, C. Vera, and X. Zhang, 2012: Changes in climate extremes and their impacts on the natural physical environment. in: Managing the risks of extreme events and disasters to advance climate change adaptation. Tech. rep., Intergovernmental Panel on Climate Change (IPCC). URL [https://www.ipcc.ch/site/assets/uploads/2018/03/SREX-Chap3\\_FINAL-1.pdf](https://www.ipcc.ch/site/assets/uploads/2018/03/SREX-Chap3_FINAL-1.pdf).
- Seo, E., M.-I. Lee, J.-H. Jeong, R. D. Koster, S. D. Schubert, H.-M. Kim, D. Kim, H.-S. Kang, H.-K. Kim, C. MacLachlan, and A. A. Scaife, 2019: Impact of soil moisture initialization on boreal summer subseasonal forecasts: mid-latitude surface air temperature and heat wave events. *Climate Dynamics*, **52** (3), 1695–1709, <https://doi.org/10.1007/s00382-018-4221-4>.
- Siebert, S., and D. B. Stephenson, 2019: Forecast recalibration and multimodel combination. *Sub-Seasonal to Seasonal Prediction:*

- The Gap between Weather and Climate Forecasting*, A. Robertson, and F. Vitart, Eds., Elsevier, chap. 15, 321–336.
- Sobolowski, S., G. Gong, and M. Ting, 2010: Modeled Climate State and Dynamic Responses to Anomalous North American Snow Cover. *Journal of Climate*, **23** (3), 785–799, <https://doi.org/10.1175/2009JCLI3219.1>, publisher: American Meteorological Society Section: Journal of Climate.
- Subramanian, A. C., M. A. Balmaseda, L. Centurioni, R. Chattopadhyay, B. D. Cornuelle, C. DeMott, M. Flatau, Y. Fujii, D. Giglio, S. T. Gille, T. M. Hamill, H. Hendon, I. Hoteit, A. Kumar, J.-H. Lee, A. J. Lucas, A. Mahadevan, M. Matsueda, S. Nam, S. Paturi, S. G. Penny, A. Rydbeck, R. Sun, Y. Takaya, A. Tandon, R. E. Todd, F. Vitart, D. Yuan, and C. Zhang, 2019: Ocean observations to improve our understanding, modeling, and forecasting of subseasonal-to-seasonal variability. *Frontiers in Marine Science*, **6**, <https://doi.org/10.3389/fmars.2019.00427>.
- TheGuardian, 2018: UK summer 'wind drought' puts green revolution into reverse. URL <https://tinyurl.com/5b2m8mj8>.
- Toth, Z., and R. Buizza, 2019: Chapter 2 - weather forecasting: What sets the forecast skill horizon? *Sub-Seasonal to Seasonal Prediction*, A. W. Robertson, and F. Vitart, Eds., Elsevier, 17–45, <https://doi.org/10.1016/B978-0-12-811714-9.00002-4>.
- Unger, D. A., 1985: A method to estimate the Continuous Ranked Probability score. *Conference on Probability and Statistics in Atmospheric Sciences*, **9th**, 206–213, URL [https://jglobal.jst.go.jp/en/detail?JGLOBAL\\_ID=200902092398162270](https://jglobal.jst.go.jp/en/detail?JGLOBAL_ID=200902092398162270).
- van den Hurk, B., F. Doblas-Reyes, G. Balsamo, R. D. Koster, S. I. Seneviratne, and H. Camargo, 2012: Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe. *Climate Dynamics*, **38** (1), 349–362, <https://doi.org/10.1007/s00382-010-0956-2>.
- Van Oldenborgh, G. J., F. Doblas-Reyes, B. Wouters, and W. Hazeleger, 2012: Decadal prediction skill in a multi-model ensemble. *Climate Dynamics*, **38**, 1263–1280, <https://doi.org/10.1007/s00382-012-1313-4>.
- Vigaud, N., M. K. Tippett, J. Yuan, A. W. Robertson, and N. Acharya, 2019: Probabilistic Skill of Subseasonal Surface Temperature Forecasts over North America. *Weather and Forecasting*, **34** (6), 1789–1806, <https://doi.org/10.1175/WAF-D-19-0117.1>, publisher: American Meteorological Society Section: Weather and Forecasting.
- Vitart, F., A. W. Robertson, and D. L. Anderson, 2012: Subseasonal to seasonal prediction project: Bridging the gap between weather and climate. *Bulletin of the World Meteorological Organization*, **61** (2), 23.
- Wagner, A. J., 1989: Medium- and long-range forecasting. *Weather and Forecasting*, **4** (3), 413 – 426, [https://doi.org/10.1175/1520-0434\(1989\)004<0413:MALRF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1989)004<0413:MALRF>2.0.CO;2).
- Wilks, D. S., 2019: *Statistical methods in the atmospheric sciences*. 4th ed., Elsevier, Cambridge.
- WindEurope, 2022: Wind energy in Europe: 2021 statistics and the outlook for 2022-2026. WindEurope, Rue Belliard 40, B-1040 Brussels, URL <https://tinyurl.com/5bydexpd>.
- WMO, 2017: Guidelines for nowcasting techniques. Tech. rep., Geneva, Switzerland.

URL [https://library.wmo.int/doc\\_num.php?explnum\\_id=3795](https://library.wmo.int/doc_num.php?explnum_id=3795).

Woolnough, S. J., F. Vitart, and M. A. Balmaseda, 2007: The role of the ocean in the Madden–Julian Oscillation: Implications for MJO prediction. *Quarterly Journal of the Royal Meteorological Society*, **133** (622), 117–128, <https://doi.org/10.1002/qj.4>.

Zhang, F., Y. Q. Sun, L. Magnusson, R. Buizza, S.-J. Lin, J.-H. Chen, and K. Emanuel, 2019: What is the predictability limit of midlatitude weather? *Journal of the Atmospheric Sciences*, **76** (4), 1077 – 1091, <https://doi.org/10.1175/JAS-D-18-0269.1>.

Zheng, C., E. K.-M. Chang, H.-M. Kim, M. Zhang, and W. Wang, 2018: Impacts of the Madden–Julian Oscillation on storm-track activity, surface air temperature, and precipitation over North America. *Journal of Climate*, **31** (15), 6113 – 6134, <https://doi.org/10.1175/JCLI-D-17-0534.1>.

Zhu, H., M. C. Wheeler, A. H. Sobel, and D. Hudson, 2014: Seamless Precipitation Prediction Skill in the Tropics and Extratropics from a Global Model. *Monthly Weather Review*, **142** (4), 1556–1569, <https://doi.org/10.1175/MWR-D-13-00222.1>, publisher: American Meteorological Society Section: Monthly Weather Review.





## 2 Quantitative assessment of sub-seasonal predictions

What gets measured gets improved.

— Peter Drucker

### Key conclusions

### Objective

The main objective of this chapter is to assess the quality of European sub-seasonal predictions of 100-m wind speed and 2-m temperature and understand their predictability limits.

### Data and Methods

We use sub-seasonal forecasts and hindcasts of 100-m wind speed and 2-m temperature over Europe from the European Centre for Medium-Range Weather Forecasts (ECMWF). We use ERA5 reanalysis as observations. We first correct the systematic bias of both forecasts and hindcasts using the Mean and Variance Adjustment method. We then carry out a comprehensive assessment of sub-seasonal prediction quality of both forecasts and hindcasts using Continuous Ranked Probability Score (CRPS), Anomaly Correlation Coefficient (ACC), and Reliability Diagrams to learn about accuracy, association, reliability, resolution, and sharpness attributes.

- Temperature is generally more predictable than wind speed;
- The prediction skill is generally higher in winter compared to other seasons for both variables;
- Forecasts, due to their larger ensemble size, are more skillful than hindcasts;
- The skill horizon of weekly mean values, based on the Continuous Ranked Probability Skill Score, is about six weeks for temperature predictions, and about three weeks for wind speed predictions.

### Publication

This chapter has been published in *Monthly Weather Review* in June 2022 (©**American Meteorological Society**. **Used with permission**).

## How Skillful Are the European Subseasonal Predictions of Wind Speed and Surface Temperature?

NAVEEN GOUTHAM,<sup>a,b</sup> RIWAL PLOUGONVEN,<sup>b</sup> HIBA OMRANI,<sup>a</sup> SYLVIE PAREY,<sup>a</sup> PETER TANKOV,<sup>c</sup> ALEXIS TANTET,<sup>b</sup> PETER HITCHCOCK,<sup>d</sup> AND PHILIPPE DROBINSKI<sup>b</sup>

<sup>a</sup> EDF Laboratory Paris-Saclay, Palaiseau, France

<sup>b</sup> Laboratoire de Météorologie Dynamique-IPSL, Ecole Polytechnique, Institut Polytechnique de Paris, ENS, PSL Research University, Sorbonne Université, CNRS, France

<sup>c</sup> CREST/ENSAE, Institut Polytechnique de Paris, Palaiseau, France

<sup>d</sup> Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York

(Manuscript received 4 August 2021, in final form 3 March 2022)

**ABSTRACT:** Subseasonal forecasts of 100-m wind speed and surface temperature, if skillful, can be beneficial to the energy sector as they can be used to plan asset availability and maintenance, assess risks of extreme events, and optimally trade power on the markets. In this study, we evaluate the skill of the European Centre for Medium-Range Weather Forecasts' subseasonal predictions of 100-m wind speed and 2-m temperature. To the authors' knowledge, this assessment is the first for the 100-m wind speed, which is an essential variable of practical importance to the energy sector. The assessment is carried out on both forecasts and reforecasts over European domain gridpoint wise and also by considering several spatially averaged domains, using several metrics to assess different attributes of forecast quality. We propose a novel way of synthesizing the continuous ranked probability skill score. The results show that the skill of the forecasts and reforecasts depends on the choice of the climate variable, the period of the year, and the geographical domain. Indeed, the predictions of temperature are better than those of wind speed, with enhanced skill found for both variables in the winter relative to other seasons. The results also indicate significant differences between the skill of forecasts and reforecasts, arising mainly due to the differing ensemble sizes. Overall, depending on the choice of the geographical domain and the forecast attribute, the results show skillful predictions beyond 2 weeks, and in certain cases, up to 6 weeks for both variables, thereby encouraging their implementation in operational decision-making.

**KEYWORDS:** Europe; Subseasonal variability; Forecast verification/skill; Temperature


### 1. Introduction

Subseasonal to seasonal (S2S) predictions (Vitart et al. 2017; Robertson and Vitart 2018), which refer to predictions beyond 2 weeks and up to a season, are influenced by both atmospheric initial conditions and boundary forcings (Hoskins 2012). Issuing skillful predictions on S2S time scale used to be considered difficult as it was thought that this time scale was both too long for the memory in the initial conditions to persist and too short for the changes in the boundary conditions to have a significant impact (Molteni et al. 1986; Robertson and Vitart 2018). However, recent studies (Hoskins 2012; Robertson and Vitart 2018) have shown otherwise by identifying the key sources of predictability on S2S time scale, which are the Madden–Julian oscillation (MJO) (e.g., Jones et al. 2004a,b), soil moisture (e.g., Koster et al. 2011; van den Hurk et al. 2012), snow cover (e.g., Sobolowski et al. 2010; Lin and Wu 2011), stratosphere–troposphere interaction (e.g., Baldwin et al. 2003), and ocean conditions (e.g., Woolnough et al. 2007; Fu et al. 2007). Although the predictability of small-scale phenomena and intraday variations on S2S time scales

remains poor (Robertson and Vitart 2018), predictability may persist for large scale phenomena. It is thus critical to aggregate/average values on relevant spatiotemporal scales in order to extract the predictable component of the signal by filtering out motions that behave like noise (Zhu et al. 2014).

In practice, both weather and seasonal predictions are carried out from imperfect initial conditions using imperfect numerical models (Robertson and Vitart 2018). S2S predictions fall beyond the theoretical limit of deterministic predictability (i.e., 10 days) (Lorenz 1965; Jifan 1989; Zhang et al. 2019), and hence these forecasts are produced using ensembles of numerical integrations: a future state of the atmosphere is then a range of possibilities. This transition from deterministic to probabilistic approach has been a major breakthrough in extending the skill horizon of S2S forecasts (Palmer 2012).

A continuously growing share of renewable power systems in the energy mix (International Energy Agency 2020), and changing frequency and intensity of extreme events in the form of storms, heat waves, and cold spells (Seneviratne et al. 2012) make the energy sector one of the most prominent potential end-users of S2S forecasts (White et al. 2017). The energy industry can greatly benefit from skillful S2S forecasts of geophysical variables as they can be used to plan asset availability and maintenance, assess and allocate risks of extreme events on production and consumption several weeks in advance in the framework of the “Ready-Set-Go!” approach

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Naveen Goutham, naveen.goutham@edf.fr or naveen.goutham@polytechnique.edu

DOI: 10.1175/MWR-D-21-0207.1

18

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

(White et al. 2017), improve grid efficiency, and optimally trade power on the markets. In the recent years, several studies have been conducted to assess the skill of S2S forecasts: Lynch et al. (2014) evaluated the skill of the European Centre for Medium-Range Weather Forecasts' (ECMWF) extended-range forecasts of 10-m wind speed between 2008 and 2013 for the winter months at the weekly time scale over Europe, and they found statistically significant skill beyond 14 days. Lledó and Doblas-Reyes (2020) assessed the impacts of strong MJO events on 10-m wind speed over Europe and developed a hybrid statistical–dynamical model to better predict 10-m wind speed conditioned on the MJO status. Büeler et al. (2020) studied windows of opportunity to have enhanced skill for ECMWF's S2S predictions of country and month-ahead-averaged quantities of 10-m wind speed, 2-m temperature, and precipitation following anomalous stratospheric polar vortex (SPV) events, and they found enhanced/reduced skill over certain regions in Europe following strong SPV events. Monhart et al. (2018) assessed the skill of the ECMWF's subseasonal forecasts of surface temperature and precipitation against several ground based-station data across Europe and found higher skill for temperature forecasts as compared with precipitation forecasts. They also demonstrated that the skill of temperature across Europe shows a seasonal pattern with higher skill observed in winter relative to other seasons, and a spatial pattern with improved skill observed in northern Europe. Vigaud et al. (2019) evaluated the skill of the surface temperature predictions from several forecasting systems over North America, and found skillful predictions beyond 2 weeks. Diro and Lin (2020) assessed the skill of the S2S forecasts of snow water equivalent and surface temperature from several models within the subseasonal experiment project. They also built a link between the two variables concluding that the weak snow–temperature coupling strength in the models is one of the contributing factors for lower skill of temperature forecasts. Dorrington et al. (2020) quantified the skill of S2S forecasts of surface temperature averaged across France from an end-user perspective, and emphasized basing the assessment of forecasts keeping potential end-user applications in mind.

For the energy sector, the 100-m wind speed forecasts are crucial to estimate the energy extracted from the wind (Jourdiér 2015). Nevertheless, as per our knowledge, there is no published peer-reviewed work on the assessment of S2S 100-m wind speed forecast skill. Because the 100-m wind speed is both closer to the turbine hub height and better represented in the ECMWF model relative to the 10-m wind speed (Alonzo et al. 2018), and since the vertical extrapolation of wind speed from 10 m to the turbine hub height could lead to significant errors (Jourdiér 2015), it is important to assess the skill of S2S forecasts of 100-m wind speed and understand their predictability limits. A vast majority of the published peer-reviewed research to date on the assessment of S2S surface temperature forecast skill are either limited to some ground based stations (e.g., Monhart et al. 2018) or a specific geographic domain (e.g., Vigaud et al. 2019) or restricted by a single metric (e.g., Diro and Lin 2020). In addition, the fast pace of change and improvement of S2S prediction systems is such that it is necessary to regularly revisit and update the assessment of their skill (Vitart 2014).

Although we agree with Dorrington et al. (2020) on the need and value of assessing forecasts based on end-user applications, it is also useful and complementary to assess the skill of the S2S forecasts of purely meteorological variables: this provides a baseline measure of the general skill of the forecasts, indicative independently of specific applications. This serves as a reference for further attempts to improve forecasts.

This study examines the skill of the S2S forecasts and reforecasts [note that for brevity “(re)forecasts” will be used hereinafter to indicate “forecasts and reforecasts” when referring to both at once] of 100-m wind speed and 2-m temperature at the weekly time scale in the recent versions of the ECMWF's S2S prediction system to understand the differences of skill that may arise due to differing ensemble sizes between the forecasts and the reforecasts. The assessment is carried out systematically across the European domain gridpoint wise and also by considering several spatially averaged country-sized domains to identify geographical regions with enhanced/reduced skill, using several metrics (Coelho et al. 2018) for providing a comprehensive overview of the forecast quality. The seasonal cycle of skill is also investigated in the study. The article is organized as follows: section 2 outlines the data used, section 3 describes in detail the method employed to evaluate the skill of (re)forecasts, section 4 explains the results obtained, and section 5 discusses the key findings and provides concluding remarks.

## 2. Data

### a. Forecasts and hindcasts

The operational S2S predictions (Vitart et al. 2017) from the ECMWF model (Vitart et al. 2019) are produced by extending the medium range forecasts (i.e., up to 2 weeks) to 46 days 2 times per week (at 0000 UTC on Mondays and Thursdays). These are ensemble predictions resulting from coupled ocean–atmosphere integrations. The ensemble is composed of 51 members (50 perturbed + control) obtained using singular vectors (Leutbecher 2005). Model uncertainty is represented through the stochastically perturbed parameterization tendencies scheme (Buizza et al. 1999; Palmer et al. 2009). These predictions are originally produced at a horizontal resolution of Tco639L91 (~18 km) up to a lead time of 15 days, and Tco319L91 (~36 km) thereafter (Robertson and Vitart 2018).

As a result of the imperfect representation of the physical processes and the inherent atmospheric unpredictability (Zhang et al. 2019; Žagar and Szunyogh 2020) in the prediction models, these models tend to drift significantly from the reality after approximately 1 week (or at most 10 days) of integrations. This drift needs to be corrected to obtain the maximum value out of forecasts. To do this, ECMWF produces a set of 20 hindcasts (or reforecasts) with 11 (10 perturbed + control) ensemble members each. These hindcasts are initialized using ERA5 reanalysis, and are issued for each of the past 20 years starting from the same date as the operational forecast. To illustrate, if the operational ensemble forecast with 51 members is initiated on 14 February 2019, the hindcast

set consists of 11 ensemble members starting on 14 February 1999, 14 February 2000, ..., 14 February 2018 (Fig. 1). This hindcast set with 220 integrations (20 years  $\times$  11 members) allows us to evaluate the model climatology, and is used to calibrate the operational version.

For the purpose of this study, only the perturbed members (refer data statement to learn about the missing control) of forecasts and their corresponding hindcasts of temperature at 2 m, zonal and meridional components of 100-m wind are retrieved between December 2015 and November 2019 at a temporal resolution of 6 h (i.e., instantaneous values at 0000, 0600, 1200, and 1800 UTC) and a spatial resolution of  $0.9^\circ$  over Europe ( $34^\circ$ – $74^\circ$ N,  $13^\circ$ W– $40^\circ$ E). The forecasts and hindcasts data are retrieved from the ECMWF's Meteorological Archival and Retrieval System (MARS). The 100-m wind speed is computed as the square root of the sum of the squares of the zonal and meridional components. The forecast model underwent several cycles of improvement during the same period, and hence, the dataset consists of forecasts and hindcasts from the versions CY41R1 (1 December 2015–7 March 2016), CY41R2 (8 March 2016–21 November 2016), CY43R1 (22 November 2016–10 July 2017), CY43R3 (11 July 2017–5 June 2018), CY45R1 (6 June 2018–10 June 2019), and CY46R1 (11 June 2019–30 November 2019). Only two of them included important changes: CY41R2 benefited from an increased atmospheric resolution, whereas CY43R1 included an increase in the oceanic resolution and the addition of dynamic sea ice. Nevertheless, the differences in statistics between the different versions of the model are marginal (see appendix A).

### b. Reference

Ideally, it is preferred to verify the quality of forecasts against observations. In the absence of a serially complete, spatially coherent observed dataset, it is a common practice to verify the forecasts against reanalysis (Kalnay 2003). ECMWF produces its own fifth-generation high-resolution (1 h; 31 km) reanalysis named ERA5 using 4D-Var data assimilation and the CY41R2 cycle of the Integrated Forecast System (Hersbach et al. 2020). In this work, ERA5 data of temperature at 2 m and zonal and meridional components of wind at 100 m are retrieved from January 1979 to 2020 over Europe ( $34^\circ$ – $74^\circ$ N,  $13^\circ$ W– $40^\circ$ E) at a spatial resolution of  $0.25^\circ$  and a temporal resolution of 6 h (i.e., instantaneous values at 0000, 0600, 1200, and 1800 UTC) from the Copernicus Climate Change Services' Climate Data Store (Raoult et al. 2017). The data are then regridded to  $0.9^\circ$  using bilinear interpolation (Cionni et al. 2018, 14–21) to match the resolution of the forecasts/hindcasts. The choice of the method is motivated by the fact that ECMWF uses bilinear interpolation as the default method for the interpolation of continuous variables of ERA5 data (<https://tinyurl.com/v3d47mw8>). The 100-m wind speed is computed from the wind components as previously described. Despite the biases, ERA5 reanalysis represents the surface wind speed (e.g., Ramon et al. 2019; Jourdiere 2020; Brune et al. 2021) and surface temperature (e.g., Simmons et al. 2021) well, with small errors that are acceptable for

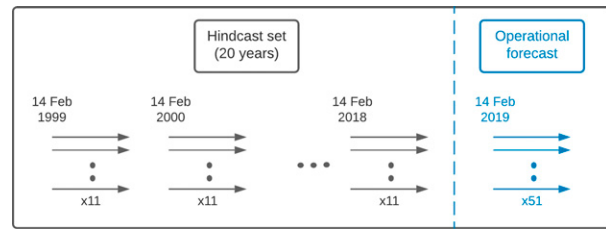


FIG. 1. Illustration of the hindcast set for the operational forecast issued on 14 Feb 2019. The ensemble size is indicated below the arrow.

verification purposes. Consequently, the ERA5 reanalysis data of 2-m temperature and 100-m wind speed act as reference/truth in this study against which the forecasts and hindcasts are verified.

### 3. Method

The forecasts data under consideration span and represent only 4 years of climatic variability, i.e., from December 2015 to November 2019. Although the forecasts, initialized using operational analysis with ensemble size 5 times as large as the hindcasts, are expected to better represent uncertainty in initial conditions and predictions, conclusions obtained from the verification of forecasts alone may be misleading because the climate during this period of time may have been more (or less) favorable for skillful predictions (Jung et al. 2011). In contrast, hindcasts, spanning 23 years from December 1995 to November 2018, represent climate variability that is 6 times as long as that of the forecasts, and can be used to perform a robust model skill assessment. However, reforecasts are likely less reliable because the ensemble size is smaller by a factor of 5 relative to forecasts and because of the way they are initialized. Hence, both forecasts and reforecasts are assessed in this study so as to understand the skill differences.

In the absence of reliable forecasts, for end-users or for different applications, a common practice is to use observed climatology, a long term average estimated from available historical observed data for the area and time period of the year concerned, as the expected weather. Therefore, it is often encouraged to not just assess the quality of forecasts, but also their relative value with respect to observed climatology. In this work, the observed climatology for each of the evaluated forecasts in any given time of the year is constructed from ERA5 reanalysis by taking the values of each of the past 35 years for the same time period of the year under consideration. To illustrate, for the forecasts issued on 14 February 2019, the observed climatology consists of weekly averaged ERA5 data starting on 14 February 1984, 14 February 1985, ..., 14 February 2018. This also implies that the forecasts issued in 2015 and 2018 have different observed climatology (i.e., rolling climatology) in order to take into account the climatic trend. However, because of the limited availability of ERA5 (i.e., from January 1979 onward at the time of commencement of this study), each of the reforecasts within a given hindcast set has the

same observed climatology as the corresponding forecast (e.g., all the hindcasts demonstrated in Fig. 1 have the same observed climatology as that of the forecast of 14 February 2019). The choice of observed climatology for the reforecasts may have a consequence on the skill of the reforecasts in a sense that the climatology may be favored over reforecasts while computing the skill score in the case of an extreme event because the information about the event is already present in the observed climatology. Nevertheless, the likelihood of witnessing such events is very low. Anyhow, this problem could be averted by using the back-extended ERA5 (i.e., from 1950 onward) and constructing climatology for the reforecasts in the same way as that of the forecasts.

#### a. Bias adjustment of (re)forecasts

Calibration is a joint property of forecasts and observations. A probabilistic forecast is perfectly calibrated, if, when averaged over several forecasts, the forecast probabilities match the observed frequencies. In addition to the chaotic nature of the atmosphere, the lack of reliable and calibrated forecasts may arise due to either one or a combination of initialization errors, model errors, model parameterizations, truncation errors, and missing physical processes. The lack of forecast calibration could be observed in the form of forecast mean bias, dispersion error, lack of association with reality, lack of reliability, or imperfect representations of trend and variability, to quote a few. If the forecast is well calibrated, the average error in the ensemble mean should be indicative of the ensemble spread, and the variance of the forecast model climatology should be equivalent to that of the climatological truth (Wilks 2019). There exist several methods with varying levels of sophistication to calibrate forecasts (Manzanas et al. 2019). Manzanas et al. (2019) have shown that simple bias adjustment methods such as mean and variance adjustment (MVA) can perform as well as the sophisticated calibration techniques such as nonhomogeneous Gaussian regression in correcting model biases. Their study also highlighted that the additional value gained by using sophisticated calibration techniques over simple bias adjustment methods are only marginal and are limited to certain geographical regions (e.g., tropics) and/or seasons. In this study, the bias adjustment of the (re)forecasts is carried out using the MVA method as described in Leung et al. (1999), Torralba et al. (2017), and Manzanas et al. (2019). The bias adjusted ensemble member  $j$  of any forecast at any given lead time is given by

$$x_j^* = (x_j - \bar{x}_e) \frac{\sigma_{\text{ref}}}{\sigma_e} + \bar{o}_{\text{ref}}, \quad (1)$$

where  $x_j$  is the member whose bias needs to be adjusted;  $\bar{x}_e$  and  $\sigma_e$  are the mean and the standard deviation, respectively, of all the members of all the hindcasts corresponding to the forecast; and  $\bar{o}_{\text{ref}}$  and  $\sigma_{\text{ref}}$  are the mean and the standard deviation, respectively, of the truth (or observations) corresponding to the hindcasts. For the bias adjustment of any given reforecast within a hindcast set, the remaining 19 years of hindcasts are used to adjust the mean and the

spread through a leave-one-out approach to prevent overfitting.

#### b. Measures of predictive skill

The predictive skill of a point forecast could be evaluated by measuring the correspondence between the forecast and the observation through simple scores such as the mean absolute error, the mean squared error, or the root mean squared error (Jolliffe and Stephenson 2003). To assess the skill of the probabilistic forecasts, several scores have been proposed in the literature, each one assessing a specific attribute of forecast quality (Wilks 2019; Coelho et al. 2019). One of the important and most widely used scores to evaluate the skill of the full predictive distribution of the probabilistic forecasts of continuous predictands is the continuous ranked probability score (CRPS) (Matheson and Winkler 1976; Unger 1985; Hersbach 2000). The CRPS is the area under the curve that is formed by computing the squared difference between the cumulative distribution functions (CDFs) of the forecast and the observation. When the observation is a single number, as is often the case, its CDF is a Heaviside step function centered on that value. This implies that if the forecast is deterministic, CRPS simplifies to the absolute error between the forecast and the observation:

$$\text{CRPS} = \int_{-\infty}^{\infty} [F(y) - F_e(y)]^2 dy, \quad (2)$$

where  $F(y)$  is the empirical CDF of forecasts/(re)forecasts computed by taking weekly mean of each of the ensemble members and

$$F_e(y) = \begin{cases} 0, & \text{if } y < e \\ 1, & \text{if } y \geq e \end{cases} \quad (3)$$

is the CDF of observation for the observed weekly mean  $e$ , denoted as a step function that jumps from 0 to 1 at the point where the forecast is equal to the observation.

The CRPS is negatively oriented (i.e., smallest values indicate more accurate forecasts), and it rewards those forecasts whose probabilities are concentrated around the observation. As the lead time increases, the ability to predict finer-scale features in time and space quickly diminishes. Consequently, in this work, the CRPS is computed for weekly averaged quantities by considering (re)forecasts and observations depending on the lead time and start date [Eq. (2)]. Spatial averaging, wherever applicable, is performed by taking the mean of cosine-latitude weighted gridpoint values in order to obtain a single scalar time series over the domain, for which all the metrics are computed (Dorrington et al. 2020). The CRPS has the same units as the physical quantity being assessed. The CRPS can also be calculated for the observed climatology: the CDF is then obtained from the weekly means of the targeted time period of the year in each of the years covered.

The relative value of the forecasts/(re)forecasts with respect to climatology is measured using the continuous ranked probability skill score (CRPSS) as described in Eq. (4). It can be



observed from Eq. (4) that skillful (re)forecasts should have a CRPSS  $> 0$ . The CRPSS is bounded above by 1, but negative values are unbounded;

$$\text{CRPSS} = 1 - \frac{\text{CRPS}_{(\text{re})\text{forecasts}}}{\text{CRPS}_{\text{climatology}}}. \quad (4)$$

The standard practice in forecast verification is to compute the average CRPS of several reforecasts at each of the lead times considered and compare it with the average CRPS of climatology to obtain an average CRPSS. The averaging is done in order to assess the reliability component of forecasts (Hersbach 2000), which can only be assessed with multiple forecast instances. However, as we show in section 4a, averaging ratios (i.e., CRPSS) overemphasizes negative instances, and can therefore be misleading in evaluating forecast skill. Besides, within the framework of the S2S forecasts, Coelho et al. (2019) recommended to use novel verification metrics that are meaningful to the end-users. Keeping in mind the S2S end-users in the energy sector, we propose “proportion of skillful (re)forecasts” as a novel way of synthesizing the CRPSS to measure skillfulness of the (re)forecasts. As the name suggests, the proportion of skillful (re)forecasts measures the proportion of (re)forecasts that have CRPS lower than that of climatology. To compute the proportion of skillful (re)forecasts [Eq. (5)], we first compute CRPSS of each of the (re)forecasts separately at each of the specified lead times, and then compute at each of the specified lead times the ratio of the number of (re)forecasts with CRPSS greater than zero to the total number of (re)forecasts. Accordingly, skillful (re)forecasts should have values  $> 50\%$ . In general, the proportion of skillful (re)forecasts is consistent with the median of the CRPSS of (re)forecasts. The proportion of skillful (re)forecasts is flexible in the sense that the threshold of “CRPSS greater than zero” can be adjusted (i.e., increased), in particular in situations where it would be useful to have CRPSS above a given threshold, not just above zero. The 95% confidence intervals for the proportion of skillful (re)forecasts in this study are computed using the standard parametric approach by assuming a normal distribution for the underlying data (Machin et al. 2013):

$$\begin{aligned} & \text{proportion of skillful (re)forecasts} \\ &= \frac{\text{no. of (re)forecasts with CRPSS} > 0}{\text{total no. of (re)forecasts}} \times 100. \quad (5) \end{aligned}$$

Assessing the skill of probabilistic forecasts generally involves assessing several forecast attributes (Coelho et al. 2019). Another commonly used score is the anomaly correlation coefficient (ACC), which measures the linear association between the ensemble mean and the observations. ACC is a deterministic score that is computed as the usual Pearson’s correlation between forecast ensemble mean and observed anomaly pairs of several independent forecasts (Namias 1952). The ACC for weekly averaged quantities of  $n$  independent (re)forecasts at any given place or across any given domain is given by

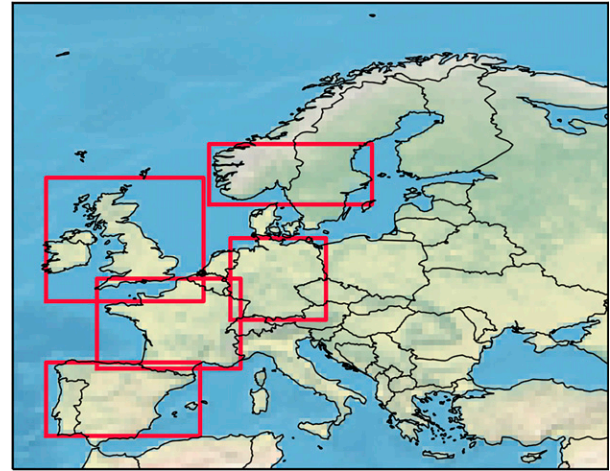


FIG. 2. Illustration of the five countrywide domains and the European domain considered in this study.

$$\text{ACC} = \frac{\text{covariance}(y', o')}{\sigma_{y'} \sigma_{o'}}, \quad (6)$$

where  $y'$  is the (re)forecast anomaly computed by removing the weekly mean climatology from the bias adjusted (re)forecast weekly ensemble mean,  $o'$  is the observed anomaly computed by removing the weekly mean climatology from the observed weekly mean, and  $\sigma$  is the standard deviation. The 95% confidence intervals for the ACC, wherever applicable, are computed through a nonparametric bootstrap approach carried out 1000 times. In general, for positioning of the synoptic scale features, the skillful forecasts should have  $\text{ACC} > 60\%$ , below which the value in the forecasts becomes only marginally useful (Robertson and Vitart 2018). Other important attributes of forecast quality such as reliability (i.e., a measure of calibration of the issued forecast probabilities), resolution (i.e., a measure of how the frequency of occurrence of the event varies as the issued forecast probability changes), and sharpness (i.e., a measure of the ability of the forecasts to produce concentrated predictive distributions that are different from the climatological probabilities) are also assessed for the weekly mean tercile, quartile, and decile forecasts through the use of reliability diagrams (Sanders 1963; Jolliffe and Stephenson 2003; Wilks 2019), which are discussed alongside the results in section 4b.

#### 4. Results

The first part of this section presents the general skill assessment of 2-m temperature and 100-m wind speed reforecasts averaged across the European domain (Fig. 2). Subsequently, the countrywide assessments of reforecast skill are also carried out. Moreover, the reforecast skill assessment at the gridpoint scale is investigated to explore geographical variations of skill. In the second part of this section, we compare the skill of reforecasts with the skill of forecasts to understand the skill differences considering different ensemble sizes.

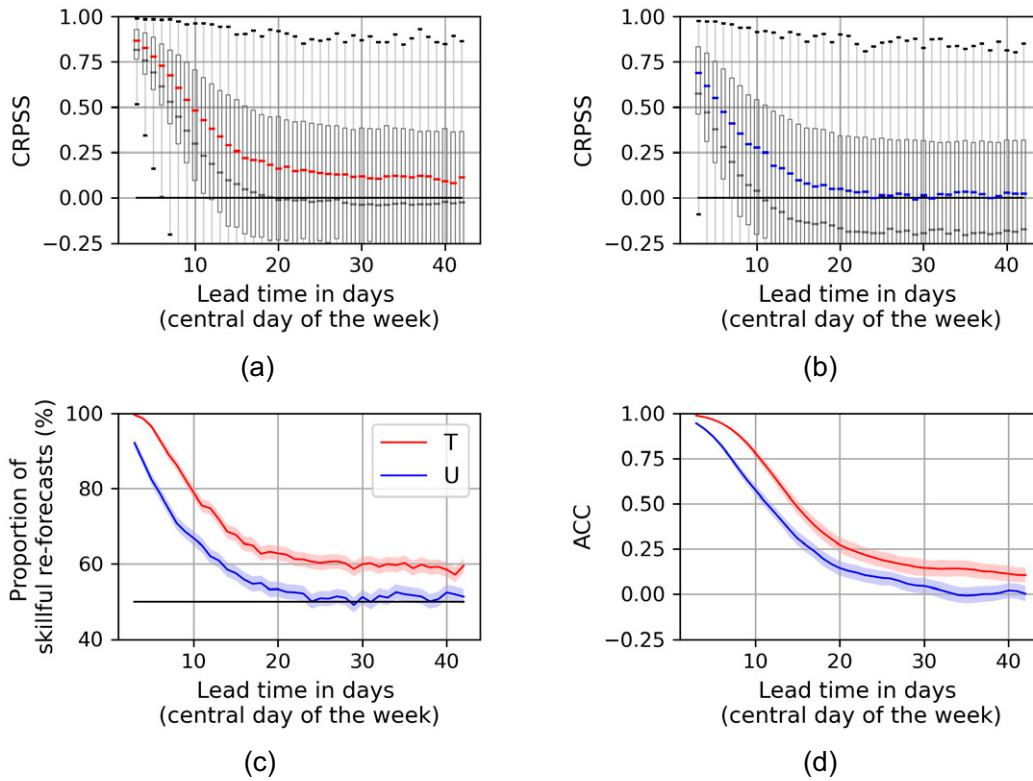


FIG. 3. Reforecast quality assessment averaged across the European domain ( $34^{\circ}$ – $74^{\circ}$ N,  $13^{\circ}$ W– $40^{\circ}$ E), showing the temporal evolution of CRPSS for (a) 2-m temperature  $T$  and (b) 100-m wind speed  $U$ , demonstrated as standard boxplots with colored bars indicating the median, gray bars indicating the mean, the gray box indicating the first and the third quartiles, whiskers indicating the end points, and outliers hidden. Values above 0 indicate skillfulness. Lead time is indicated as central day of the week (as an illustration, day 10 corresponds to the week between days 7 and 13). (c) The temporal evolution of the proportion of skillful reforecasts for the same variables. Values above 50% (black horizontal line) indicate skillfulness. (d) The temporal evolution of ACC for the same variables. Shaded regions in (c) and (d) correspond to the 95% confidence intervals.

Last, we assess reliability, resolution, and sharpness of the forecasts through the aid of reliability diagrams.

a. Reforecast skill assessments

1) GENERAL ASSESSMENT OVER EUROPE

Figure 3 compares the temporal evolution of 2-m temperature and 100-m wind speed reforecast skill for weekly mean values averaged across the European domain (Fig. 2). Figures 3a and 3b show that the mean values (i.e., gray bars) of CRPSS drop below zero after 18 and 10 days, respectively, for temperature and wind speed. On the other hand, the median (i.e., colored bars) largely stays above zero throughout all the leads, although more significantly for temperature. By definition, the skill score is bounded by 1 above, but its negative values are unbounded such that the average can be sensitive to rare, strong negative values. Hence, instead of calculating the average of the CRPSS, we compute the proportion of skillful (re)forecasts as a measure of (re)forecast skill. Figure 3c shows the temporal evolution of the proportion of skillful reforecasts with increasing lead time.

It is conspicuous that the model performs better in predicting 2-m temperature than 100-m wind speed at all lead times. While temperature reforecasts are skillful at all lead times, wind speed reforecasts are skillful until approximately day 24. Figure 3d displaying the time evolution of ACC for temperature and wind speed confirms that temperature reforecasts are more skillful than wind speed reforecasts. The ACC for temperature falls below 0.6 around day 13, and below 0.25 around day 22, whereas the ACC for wind speed falls below 0.6 around day 8, and below 0.25 around day 17. Seasonal variations of skill are discussed in the following section.

2) SEASONAL VARIATIONS OF SKILL

Seasonal variations of temperature and wind speed reforecast skill averaged across the European domain are shown in Fig. 4. It can be noticed that the proportion of skillful reforecasts is larger in the Northern Hemisphere winter [December–February (DJF)] and summer [June–August (JJA)] than in the transition seasons for both temperature (Fig. 4a) and wind speed (Fig. 4b). The ACC for temperature (Fig. 4c) is larger in



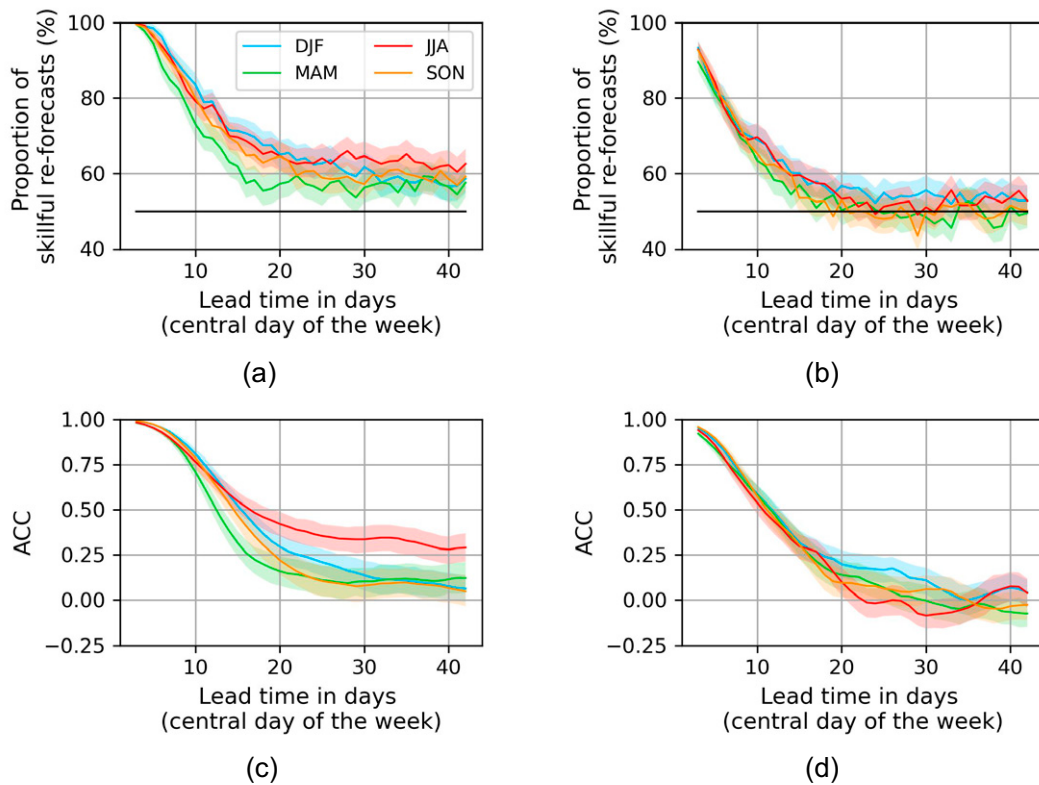


FIG. 4. Seasonal variations of reforecast quality assessment averaged across the European domain ( $34^{\circ}$ – $74^{\circ}$ N,  $13^{\circ}$ W– $40^{\circ}$ E), showing the temporal evolution of the proportion of skillful reforecasts for (a) 2-m temperature and (b) 100-m wind speed. Values above 50% (black horizontal line) indicate skillfulness. Also shown are the temporal evolution of ACC for (c) 2-m temperature and (d) 100-m wind speed. Shaded regions correspond to the 95% confidence intervals.

the summer relative to other seasons after a lead time of around 17 days. In contrast, the ACC for wind speed is larger in the winter relative to other seasons between 18 and 31 days (Fig. 4d). The improved skill in winter and/or summer may arise from stronger boundary conditions (e.g., sea surface temperature gradients), reinforced coupling (e.g., troposphere–stratosphere), enhanced memory of the initial conditions (e.g., soil moisture) among others (Robertson and Vitart 2018). However, addressing the reasons for enhanced skill in certain seasons is beyond the scope of this study. The seasons of winter and summer, in addition to demonstrating enhanced skill, also have a significant impact on the energy sector in the form of increased demand driven by heating and cooling, respectively. Therefore, only the results corresponding to winter and summer will be shown in the following sections.

### 3) COUNTRYWIDE SKILL ASSESSMENT

The assessment of reforecast skill averaged across the European domain may have limited application. In contrast, countrywide average skill of wind speed and temperature reforecasts may be closer to the scale on which end-users may need forecasts for decision-making (e.g., transmission system operators). This section investigates variations in skill over

domains typically a thousand kilometers across, e.g., over a country like the United Kingdom. Table 1 and Fig. 2 present the domains considered in this study. The choice of the domains is not just motivated by their geography (inland versus coastal, location relative to the climatological storm tracks) offering different sampling conditions, but also by their considerable share of wind power in the energy mix (International Energy Agency 2020). Figures 5 and 6 illustrate the differences in the temporal evolution of the proportion of skillful reforecasts for a selection of domains. For temperature (Fig. 5), in both DJF and JJA, predictions over Germany (e.g., proportion > 60% up to about 27 days in DJF and about 15 days in JJA) are more skillful than predictions over France (proportion > 60% up to about 20 days in DJF and about

TABLE 1. Description of the domains.

Domain region	Lower-left bound	Upper-right bound
France	43.0°N, 5.5°W	51.0°N, 7.3°E
Germany	47.3°N, 6.4°E	54.6°N, 14.9°E
Southern Scandinavia	57.6°N, 4.5°E	63.0°N, 19.0°E
Spain and Portugal	37.0°N, 10.0°W	43.5°N, 3.7°E
United Kingdom	49.0°N, 10.0°W	60.0°N, 4.0°E

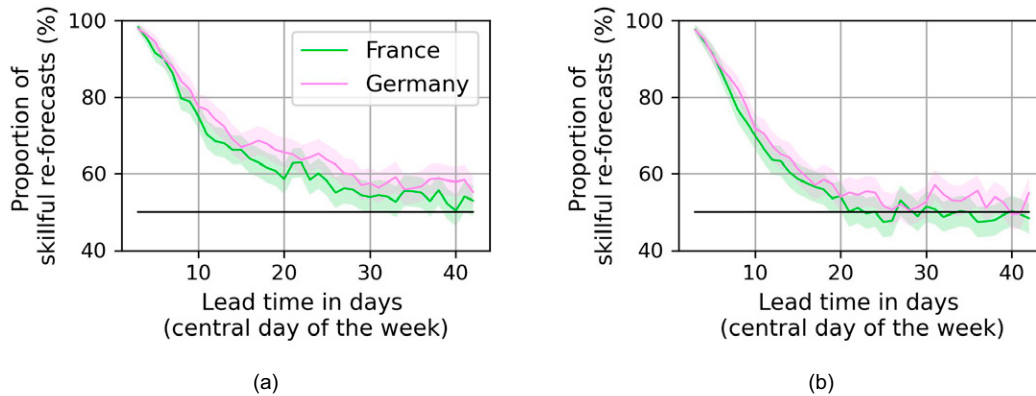


FIG. 5. Comparison of the temporal evolution of proportion of skillful 2-m temperature reforecasts between France and Germany for (a) DJF and (b) JJA. Shaded region correspond to the 95% confidence intervals. Values above 50% (black horizontal line) indicate skillfulness.

14 days in JJA). The reforecasts are more skillful in the winter as compared with the summer across both domains. The ACC for temperature also shows similar behavior to that of the proportion of skillful reforecasts (not shown). The skill of winter temperature reforecasts for Spain and Portugal and the United Kingdom (not shown) domains are similar to that for France, and the skill across southern Scandinavia (not shown) is at least as good as, if not slightly better than Germany. Overall, in DJF, the skill across the most skillful country-sized domains (e.g., Germany) is almost as good as the skill across the European domain (Fig. 4a). Whereas in JJA, the skill across the European domain is better than the skill of the most skillful country-sized domains.

The wind speed reforecasts (Fig. 6) across the United Kingdom (e.g., proportion > 60% up to about 17 days) are more skillful than that of France (proportion >60% up to about 11 days) in DJF. However, in JJA, France demonstrates marginally larger skill than the United Kingdom after about 10 days. The ACC again displays a similar pattern to that of the proportion of skillful reforecasts (not shown). In winter,

the skill across Germany and Spain and Portugal (not shown) are comparable to that of France, and the skill across southern Scandinavia (not shown) is at least as good as, if not marginally better than the United Kingdom. In DJF the skill across the United Kingdom is better than the skill across the European domain (Fig. 4b), whereas in JJA the opposite is true.

4) GRIDPOINT SKILL ASSESSMENT

Although countrywide domains are useful in predicting national averages of the variables, gridpoint assessment of skill are more appropriate to explore the geographical variations of skill. The spatial resolution of the data used in this study is about 90 km (i.e., 0.9°). This resolution is coarse enough for the S2S prediction models to still hold prediction skill on the subseasonal time scales (Buizza and Leutbecher 2015), and fine enough to be useful for a range of applications.

Figure 7 illustrates the maps of temporal evolution of the proportion of skillful reforecasts and ACC for temperature across the European domain. The maps for winter (top row of Fig. 7a) show the presence of a zonal (i.e., east-west) pattern

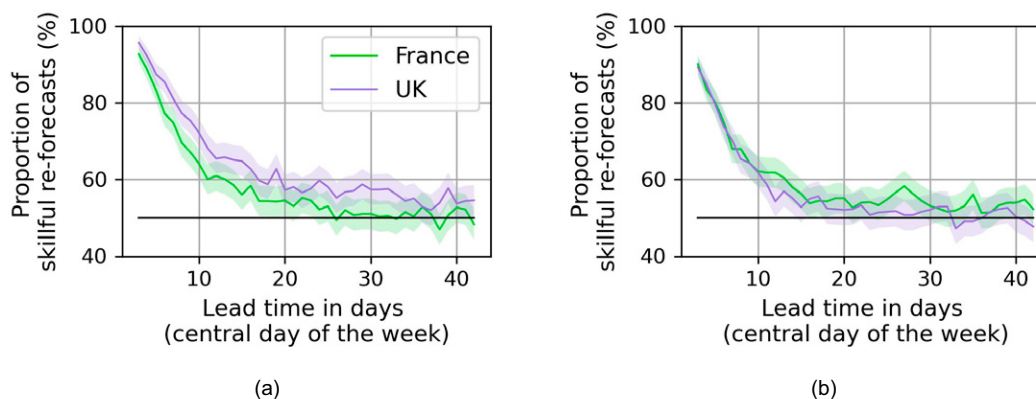
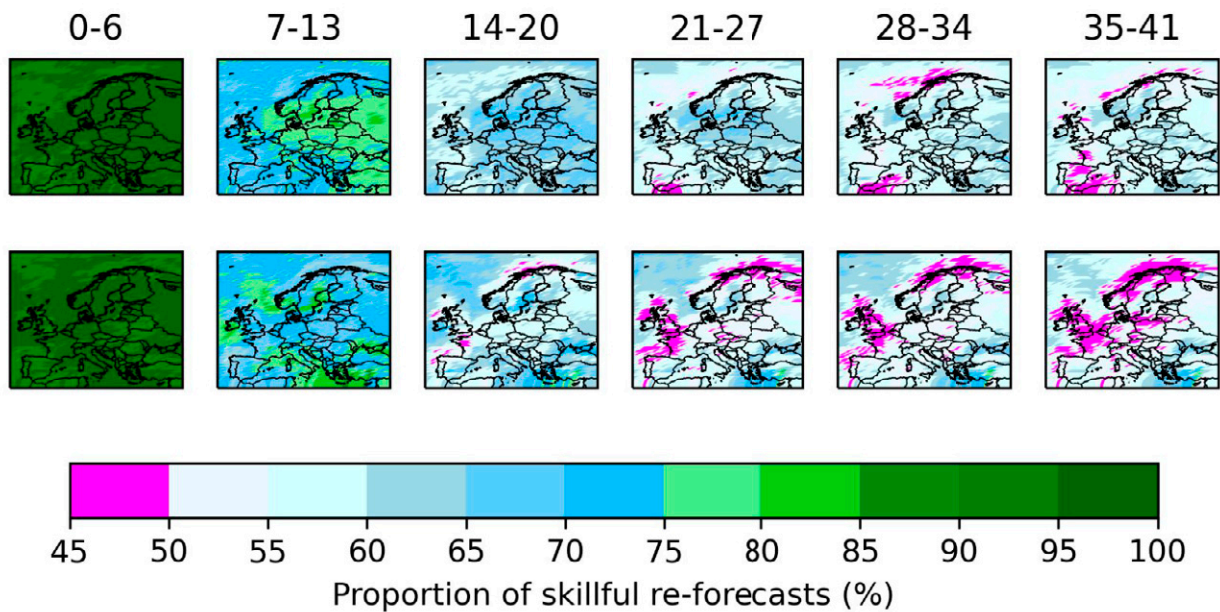
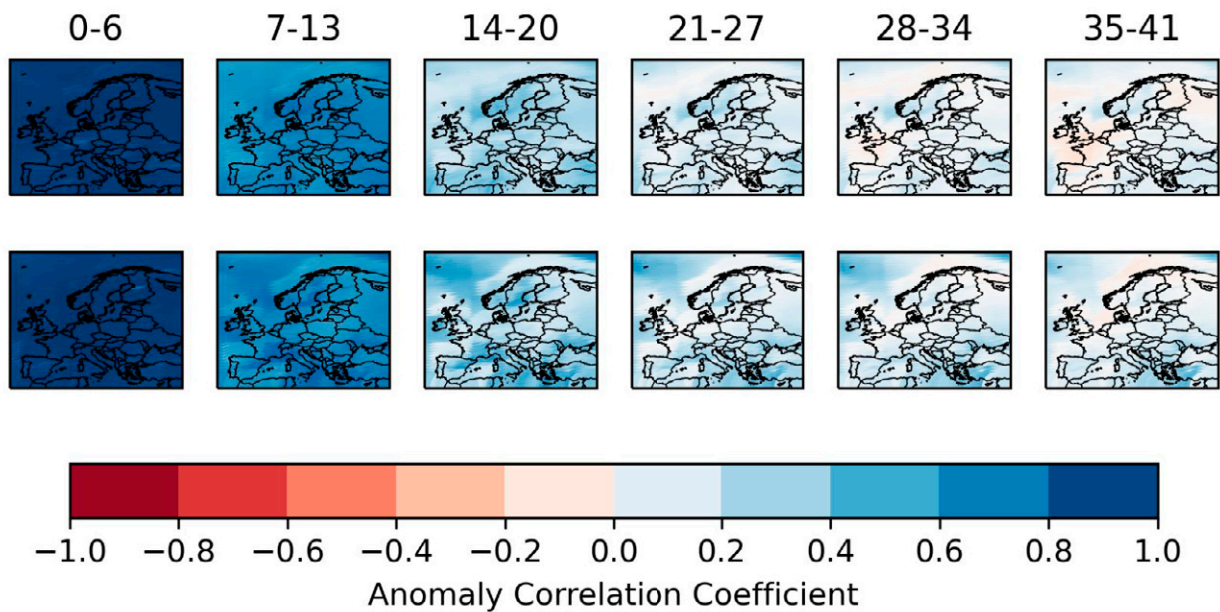


FIG. 6. Comparison of the temporal evolution of proportion of skillful 100-m wind speed reforecasts between France and the United Kingdom for (a) DJF and (b) JJA. Shaded regions correspond to the 95% confidence intervals. Values above 50% (black horizontal line) indicate skillfulness.



(a)



(b)

FIG. 7. Maps of (a) proportion of skillful reforecasts and (b) ACC for 2-m temperature over Europe. In (a) and (b), the top row is for DJF and the bottom row is for JJA. Columns from left to right show lead times centered on days 3, 10, 17, 24, 31, and 38. Values above 50% in (a) indicate skillfulness.

between second and fourth weeks indicating that temperature predictions are generally more skillful over central/eastern Europe than western Europe. In central and eastern Europe, the reforecasts are skillful even at a lead time of 6 weeks in

winter, encouraging their use in the decision-making value chain across sectors. The reforecasts are less skillful in summer in general with proportion of skillful reforecasts converging toward climatology beyond 3 weeks. The ACC (Fig. 7b)



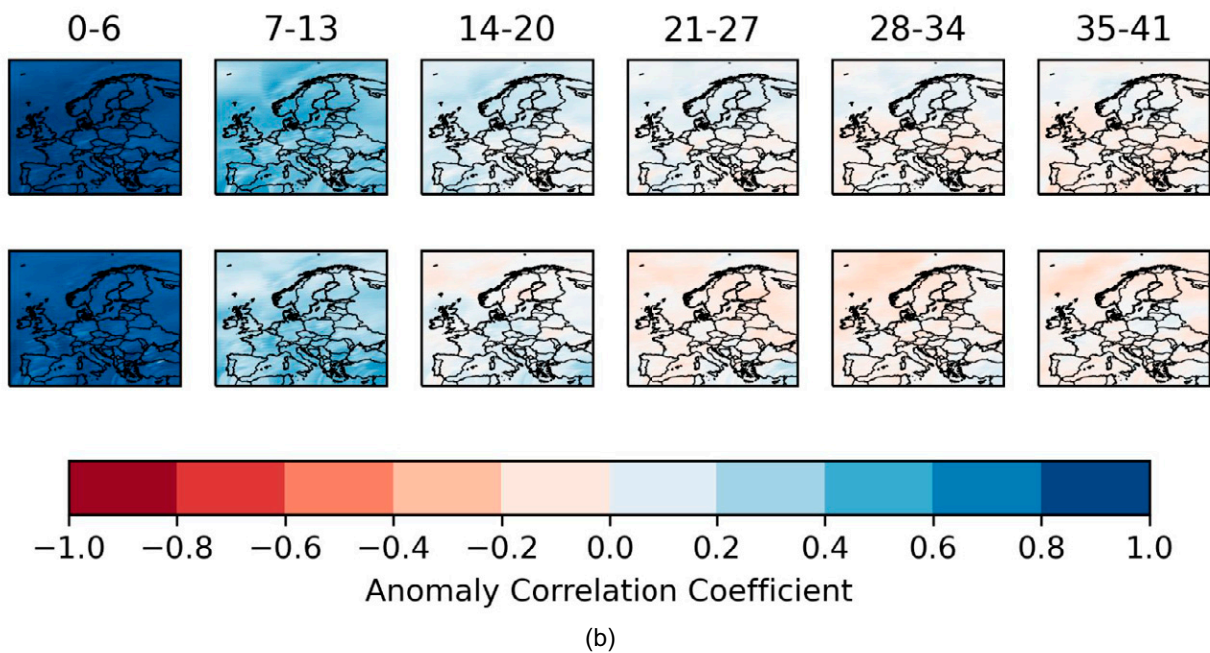
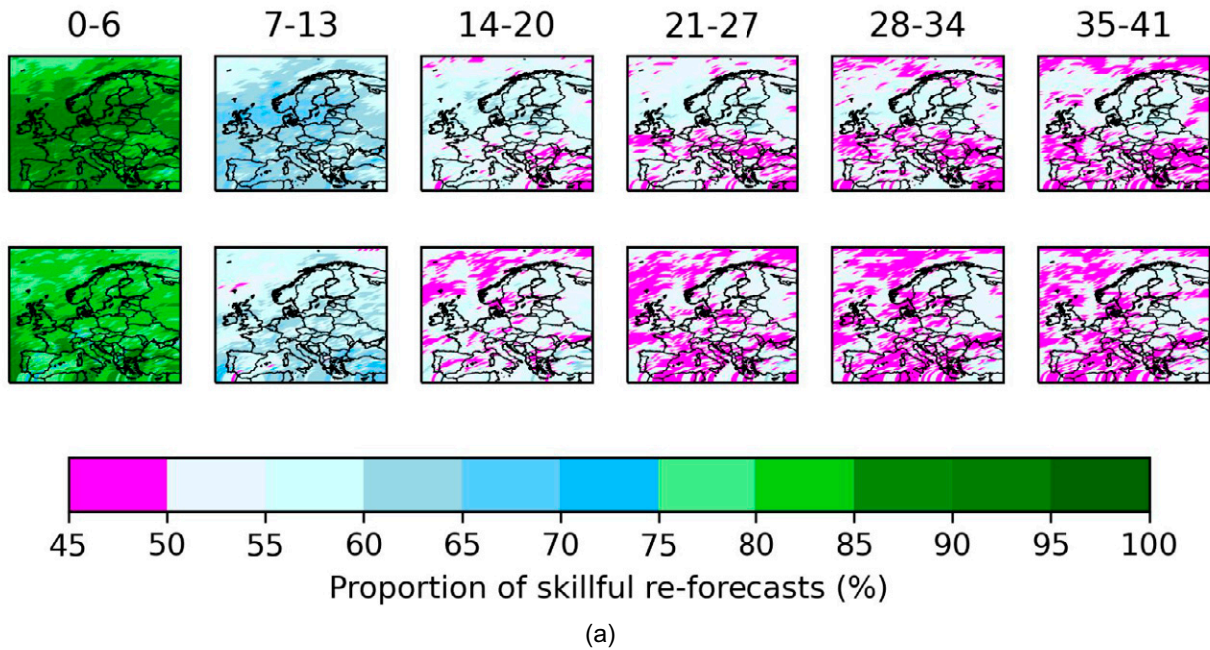


FIG. 8. As in Fig. 7, but for 100-m wind speed.

drops below 60% beyond 2 weeks showing no noticeable differences between seasons.

The maps of the proportion of skillful reforecasts and ACC for wind speed are shown in Fig. 8. Overall, the reforecasts are more skillful in winter. In addition, there exists a meridional (i.e., north–south) pattern of skill in winter (top rows in Figs. 8a,b) between second and fifth weeks indicating that

wind speed predictions are generally more skillful over northern than southern Europe. Across Scandinavia, the proportion of skillful reforecasts still exceeds 50% after 5 weeks. However, in summer, the proportion of skillful reforecasts for wind speed drops below 50% over a large part of the domain beyond 3 weeks. The ACC drops to significantly lower levels than that of temperature beyond a week in both winter and

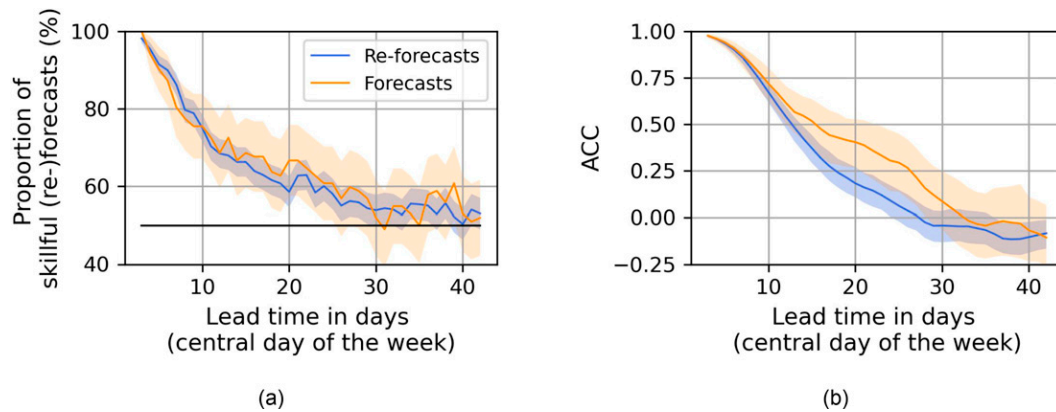


FIG. 9. Comparison of the temporal evolution of 2-m temperature forecast and reforecast skills averaged across France for DJF, showing (a) proportion of skillful reforecasts (blue) and forecasts (orange) and (b) ACC. Values above 50% (black horizontal line) indicate skillfulness. Shaded regions correspond to the 95% confidence intervals.

summer. The reasons for enhanced skill witnessed across certain regions in Figs. 7 and 8 may be related to the differences in the regional climate (i.e., maritime vs continental), low-frequency oscillations (Ardilouze et al. 2021), and to the role of circulation features, in particular the storm tracks. However, investigating the reasons for the existence of spatial pattern of skill is beyond the scope of this work.

#### b. Forecast skill assessments

The operational forecasts with 50 ensemble members each are expected to better represent uncertainty in the initial conditions and model parameterizations as compared with the reforecasts with only 10 members (Robertson and Vitart 2018). Figure 9 compares the skill between the winter forecasts and reforecasts of temperature for weekly means averaged across France (Table 1 and Fig. 2). Overall, the forecasts are more skillful than the reforecasts. The proportion of skillful forecasts (e.g., values  $> 60\%$  up to about 25 days) is essentially greater than the proportion of skillful reforecasts (values  $> 60\%$  up to about 19 days). However, the confidence intervals for the forecasts are 2 times as wide as those of the reforecasts due to a smaller sample size. The ACC of

forecasts (values  $> 0.5$  up to about 16 days, and  $> 0.25$  up to about 26 days) has a longer skill horizon when compared with that of the reforecasts (values  $> 0.5$  up to about 13 days, and  $> 0.25$  up to about 17 days). A similar pattern can also be observed with respect to other seasons and domains (not shown). The differences between the skill of the forecasts and the reforecasts of wind speed in winter are shown in Fig. 10 for the same domain. The behavior is overall comparable to that of the temperature. The significant differences between the skill of the forecasts and the reforecasts are mainly due to the differing ensemble sizes between the two (see appendix B). Even though the reforecasts (23 yr) represent a longer climatic variability than the forecasts (4 yr) and hence a better estimation of the overall skill of the model, given that they have an ensemble size that is smaller by a factor of 5 relative to the forecasts, the skill of the reforecasts should only be considered as a lower bound for the skill of the operational forecasts.

The CRPS and its respective skill score give one measure of the agreement between the forecasts and the observations. However, a thorough appreciation of the quality of forecasts requires the use of the full joint distribution of forecasts and

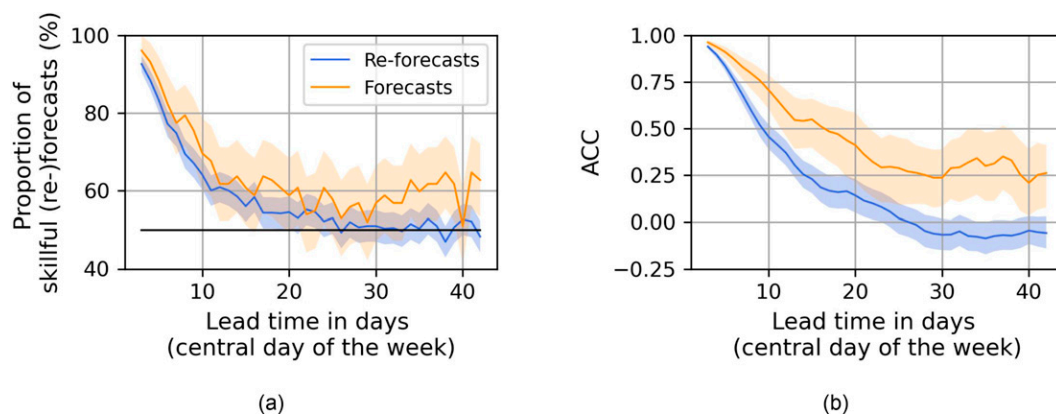


FIG. 10. As in Fig. 9, but for 100-m wind speed.

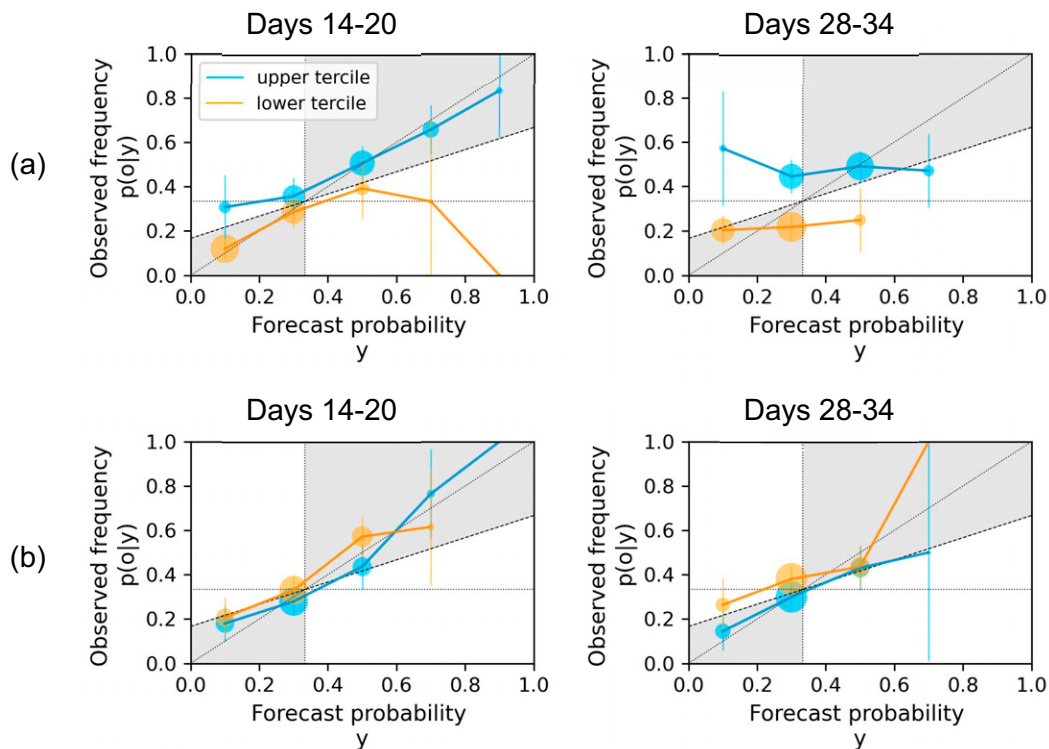


FIG. 11. Reliability diagrams for upper and lower terciles of weekly mean forecasts for the weeks centered on (left) day 17 (i.e., days 14–20) and (right) day 31 (i.e., days 28–34) averaged across France for (a) 2-m temperature and (b) 100-m wind speed. The forecasts are stratified into five bins of equal width. The size of the points is proportional to the number of forecasts in the respective bins. The vertical bars refer to the 95% confidence intervals computed through the standard parametric approach. The vertical and horizontal dotted lines indicate the climatological tercile probabilities (theoretically, the value is  $1/3$ ) in the forecasts and observations, respectively. Perfectly reliable forecasts fall on the dotted diagonal line connecting the points (0, 0) and (1, 1). The points located within the gray area contribute positively to the skill.

observations. The reliability diagram (Sanders 1963; Jolliffe and Stephenson 2003; Wilks 2019) is a graphical tool to comprehend the full joint distribution of forecasts and observations for probabilistic forecasts of a dichotomous predictand (i.e., predictand with a binary outcome). Perfectly reliable (i.e., calibrated) forecasts have observed frequencies essentially equal to forecast probabilities. Since the forecasts considered in this study have a larger ensemble size and hence a better representation of uncertainty relative to the reforecasts, the reliability diagrams are produced only for the forecasts. Figure 11 demonstrates the reliability diagrams for upper and lower terciles of weekly mean temperature and wind speed forecasts for the week centered on day 17 (i.e., days 14–20) and day 31 (i.e., days 28–34) for the France domain described in Table 1. In the figure, the lines connecting the points show no persistent offset from the 1:1 diagonal line ( $45^\circ$ ) illustrating the absence of unconditional biases. In a reliability diagram, the smaller the vertical distance between the points and the diagonal line, and the larger the vertical distance between the points and the climatological line (dotted horizontal line in the figure), the higher are the forecast reliability and resolution, respectively. Conversely, the larger

the vertical distance between the points and the diagonal line, and the smaller the distance between the points and the horizontal climatological line, the lower are the reliability and resolution, respectively. The dashed line located midway between the perfect reliability line and the horizontal climatological line represents the *no skill line*. Accordingly, the points located in the gray area contribute positively to the skill of the forecasts. For the third week of temperature forecasts (Fig. 11a), the upper-tercile forecasts are more reliable than the lower counterparts. In contrast, the reliability of the upper- and the lower-tercile wind speed forecasts are virtually comparable for both weeks (Fig. 11b). The upper tercile of temperature forecasts for the third week exhibit under-forecasting biases associated with low probabilities, and marginal over-forecasting biases associated with high probabilities. Furthermore, the lower tercile of temperature forecasts generally show significant over-forecasting associated with high probabilities, indicating poor resolution and over-confidence. For temperature in the third week, the maximum number of forecasts is located in the bins beside the climatological probability (dotted vertical line) bin (i.e., 0.2–0.4), indicating reasonable sharpness of these forecasts. In contrast, for



wind speed for both weeks, the maximum number of forecasts are concentrated in the climatological bin suggesting that the forecasts have low sharpness. For the fifth week of temperature forecasts, both the upper- and the lower-tercile forecasts are less reliable, and have a poorer resolution and sharpness as compared with the third week. While for wind speed forecasts in the fifth week, the reliability, resolution, and sharpness are comparable to that of the third week. The number of events falling within the bins are typically concentrated near low forecast probabilities. Therefore, the confidence intervals are generally narrower for lower forecast probabilities as compared with higher forecast probabilities for both upper and lower terciles. The fact that the upper-tercile temperature forecasts for the third week being more reliable and having a higher resolution than the lower-tercile forecasts, and the comparable reliability and resolution of the upper- and lower-tercile wind speed forecasts holds true for the other domains (not shown). The reliability diagrams for the upper and lower quartiles and deciles of temperature (see [appendix C](#)) and wind speed (not shown) forecasts are less reliable, and have significantly lower resolution, especially for larger forecast probabilities, relative to that of the terciles. Overall, the forecasts of temperature and wind speed carry valuable information in predicting terciles even beyond 2 weeks, encouraging their implementation in operational decision-making on this time horizon.

## 5. Conclusions

In this study, the skill of the subseasonal forecasts and reforecasts of 2-m temperature and 100-m wind speed was evaluated against ERA5 reanalysis across the European domain. The bias adjustment of the (re)forecasts was carried out using mean and variance adjustment method. To account for the different aspects of (re)forecast quality (i.e., accuracy, association, reliability, resolution, and sharpness), several metrics were applied, providing evidence that

- 1) the model generally performs better in predicting 2-m temperature than 100-m wind speed,
- 2) the skill over Europe displays a seasonal pattern with winter showing more skillful forecasts, which is followed by summer for temperature and summer/fall for wind speed,
- 3) the skill also displays a spatial pattern for temperature having more skill for eastern than for western Europe and for wind speed having more skill in northern than southern Europe,
- 4) the skill of the reforecasts should only be considered as a lower bound, and the forecasts due to their larger ensemble size represent uncertainty better and hence perform better, and
- 5) depending on the geographical domain, climate variable, and forecast attribute of choice, the weekly mean forecasts can be skillful even up to 6 weeks, encouraging their implementation in the decision-making value chain.

This study evaluated the skill of the (re)forecasts of a model originating from a single weather forecasting center (i.e.,

ECMWF). This choice was motivated by the fact that the skill of the forecasts of temperature (at 2 m and at 850 hPa) of the ECMWF model compares to or even outperforms the skill of a multimodel combination ([Hagedorn et al. 2012](#)). Nevertheless, investigation of the skill of a multimodel ensemble is an important next step. The reference data (i.e., ERA5 reanalysis) used in this study also originates from the ECMWF, produced using one of the same models (CY41R2) that is used to produce the (re)forecasts. However, verifying (re)forecasts against reanalysis produced from the same model may contribute to enhancing the skill of (re)forecasts. Hence, it is important to assess the skill of the (re)forecasts against observations or other global/regional reanalysis datasets produced using a different model. To aid in the further development of the prediction model, it is essential to understand the potential sources of predictability and the origin of model biases. The authors did not assess the skill of the (re)forecasts of other variables that are critical for the renewable energy sector such as the solar radiation and the precipitation. The authors propose to undertake these explorations in a future study.

*Acknowledgments.* Naveen Goutham thanks Association Nationale de la Recherche et de la Technologie (ANRT) for the Convention industrielle de formation par la recherche (CIFRE) fellowship of his Ph.D. This work contributes to the Energy4Climate Interdisciplinary Centre (E4C) of Institut Polytechnique de Paris and Ecole des Ponts ParisTech, supported by third Programme d'Investissements d'Avenir (ANR-18-EUR-0006-02). The authors acknowledge the data center Ensemble de services pour la recherche à l'Institut Pierre-Simon Laplace (ESPRI) for their help in storing and accessing the data. The authors thank the two anonymous reviewers who contributed significantly in improving the quality of the paper through their critical assessment.

*Data availability statement.* The archived ECMWF extended-range forecasts and hindcasts are published under a Creative Commons Attribution 4.0 International (CC BY 4.0). However, a data access fee may be applicable. Further information for the reader is available online (<https://www.ecmwf.int/>). The ECMWF ERA5 reanalysis data are publicly available and can be accessed through the Copernicus Climate Change Services' Climate Data Store upon registration. The control of the ensemble should be treated as another indistinguishable ensemble member. However, because of the unavailability of the control member in the internal database of ESPRI as a result of an unintentional human error, we had to use only the perturbed members.

## APPENDIX A

### Comparison of Skill between Different Cycles of the ECMWF Model

[Figure A1](#) shows the temporal evolution of CRPSS of reforecasts of 2-m temperature averaged across Germany ([Table 1](#)) between different cycles of the ECMWF model for four seasons

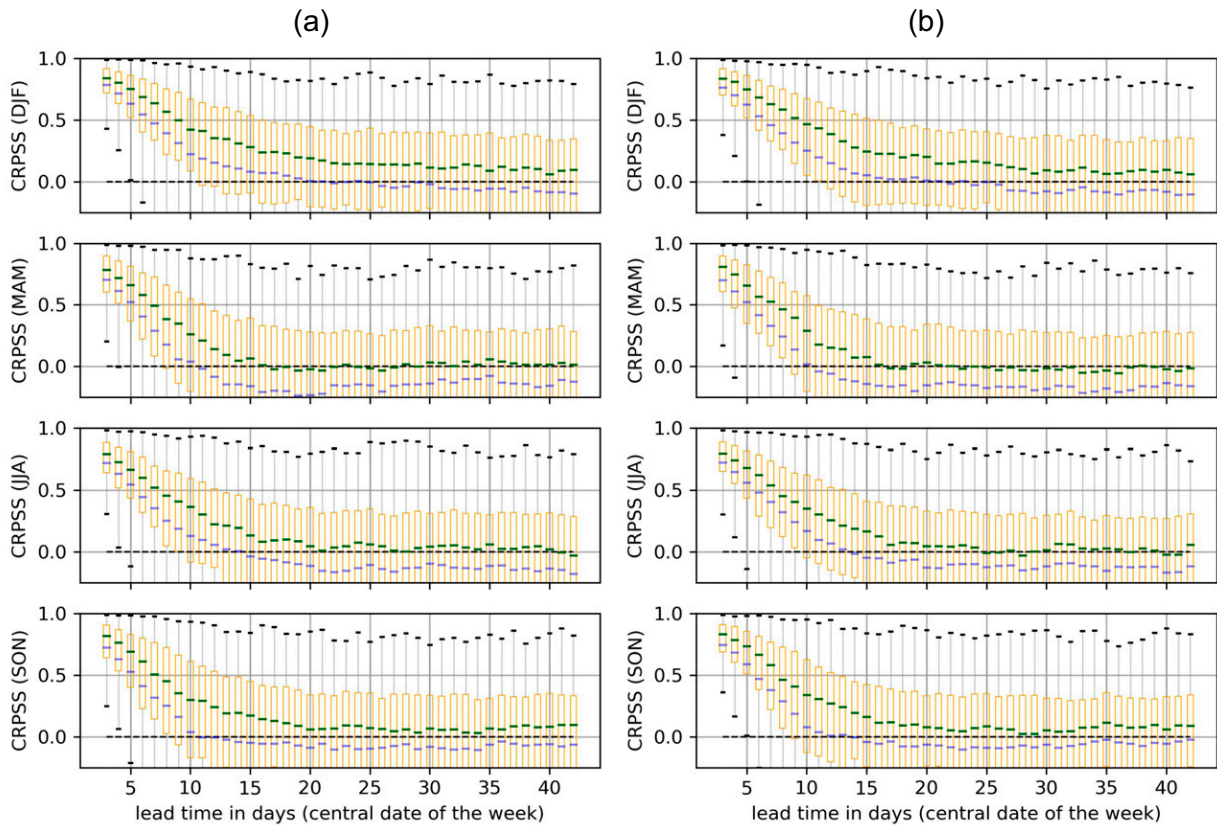


FIG. A1. Comparison of CRPSS of reforecasts of 2-m temperature averaged across Germany between different cycles of the ECMWF model for four seasons (DJF, MAM, JJA, and SON), demonstrated as standard boxplots with green bars indicating the median, blue bars indicating the mean, the orange box indicating the first and the third quartiles, whiskers indicating the end points, and outliers hidden. Values above 0 indicate skillfulness. Lead time is indicated as central day of the week (as an illustration, day 10 corresponds to the week between days 7 and 13). The reforecasts correspond to the forecasts (a) between 1 Dec 2015 and 30 Nov 2016 and (b) between 1 Dec 2018 and 30 Nov 2019.

[DJF, March–May (MAM), JJA, and September–November (SON)]. The data used in Fig. A1a consist of reforecasts mostly from the cycles CY41R1 and CY41R2, with only a few reforecasts from the cycle CY43R1. Whereas, Fig. A1b is

produced from the reforecasts data involving cycles CY45R1 and CY46R1. Although the data consists of a combination of several cycles, we can compare two versions (CY41R1 and CY46R1) by isolating DJF (top row in Fig. A1). It is very

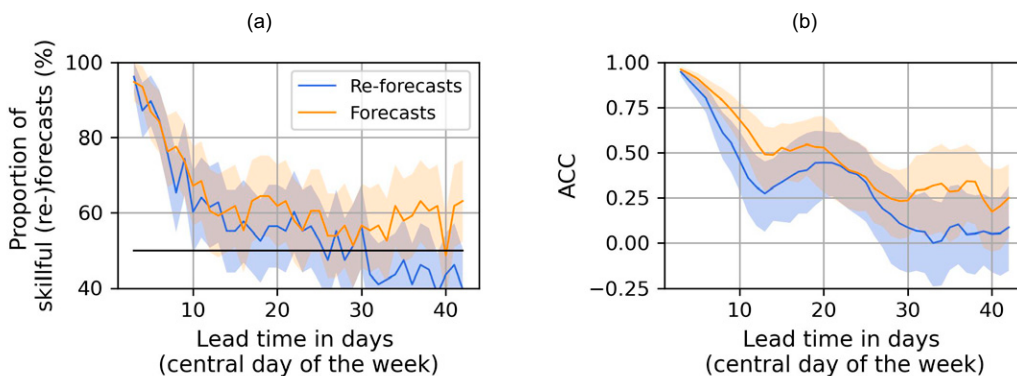


FIG. B1. Comparison of the temporal evolution of 100-m wind speed forecast and reforecast skills averaged across France for DJF between December 2015 and February 2018 (three winters), showing (a) proportion of skillful reforecasts (blue) and forecasts (orange) and (b) ACC. Shaded regions correspond to the 95% confidence intervals. In (a), values above 50% (black horizontal line) indicate skillfulness.



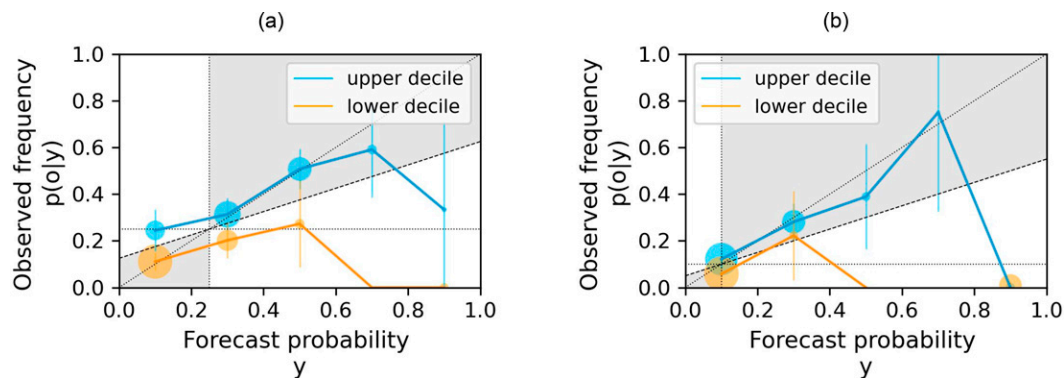


FIG. C1. Reliability diagram for upper and lower (a) quartiles and (b) deciles of weekly mean 2-m temperature forecasts for the week centered on day 17 (i.e., days 14–20) averaged across France. The forecasts are stratified into five bins of equal width. The size of the points is proportional to the number of forecasts in the respective bins. The vertical bars refer to the 95% confidence intervals computed through the standard parametric approach. The vertical and horizontal dotted lines indicate the climatological quartile/decile probabilities (theoretically, the value is  $1/4$  for a quartile and  $1/10$  for a decile) in the forecasts and observations, respectively. Perfectly reliable forecasts fall on the dotted diagonal line connecting the points (0, 0) and (1, 1). The points located within the gray area contribute positively to the skill.

difficult to say that the version CY46R1 is better than CY41R1, or vice versa. Similar observations were made with respect to other domains for both 2-m temperature and 100-m wind speed.

#### APPENDIX B

##### Comparison of Skill between Forecasts and Reforecasts

Figures 9 and 10 indicated that the forecasts are more skillful than the reforecasts. The improved skill of the forecasts may be a result of one or a combination of the way forecasts and reforecasts are initialized (forecasts are initialized using operational analysis, whereas reforecasts are initialized using ERA5 reanalysis), the difference in ensemble size (50 for forecasts and 10 for reforecasts), or the period of the sample considered (December 2015–November 2019 for forecasts and December 1995–November 2018 for reforecasts). Through this section, we try to understand the reasons for improved skill by isolating one or several factors. Figure B1 compares the temporal evolution of skill between 100-m wind speed forecasts and reforecasts similar to Figs. 9 and 10, but for the same period (i.e., DJF 2015/16, 2016/17, and 2017/18). Overall, the behavior is comparable to that of Fig. 10, with forecasts being more skillful than the reforecasts. In addition, the behavior of the ACC (Fig. B1b) of the reforecasts is similar to that of the forecasts but with lower values, indicating the importance of the role of ensemble size and the way the (re)forecasts are initialized on the skill of the (re)forecasts. In this study, since ERA5 reanalysis is used as reference against which the (re)forecasts are verified, the (re)forecast skill may not necessarily be dependent on the way (re)forecasts are initialized, thereby leaving greater weight on the ensemble size.

#### APPENDIX C

##### Reliability Diagrams for Quartiles and Deciles of Weekly Mean Temperature Forecasts

Figure C1 shows that the reliability diagrams for the upper and lower quartiles and deciles of temperature forecasts are less reliable and have significantly lower resolution, especially for larger forecast probabilities, relative to that of the terciles.

#### REFERENCES

- Alonzo, B., R. Plougonven, M. Mougeot, A. Fischer, A. Dupré, and P. Drobinski, 2018: From numerical weather prediction outputs to accurate local surface wind speed: Statistical modeling and forecasts. *Renewable Energy: Forecasting and Risk Management*, Springer Proceedings in Mathematics and Statistics, P. Drobinski et al., Eds., Springer International Publishing, 23–44, [https://doi.org/10.1007/978-3-319-99052-1\\_2](https://doi.org/10.1007/978-3-319-99052-1_2).
- Ardilouze, C., D. Specq, L. Batté, and C. Cassou, 2021: Flow dependence of wintertime subseasonal prediction skill over Europe. *Wea. Climate Dyn.*, **2**, 1033–1049, <https://doi.org/10.5194/wcd-2-1033-2021>.
- Baldwin, M. P., D. B. Stephenson, D. W. J. Thompson, T. J. Dunkerton, A. J. Charlton, and A. O'Neill, 2003: Stratospheric memory and skill of extended-range weather forecasts. *Science*, **301**, 636–640, <https://doi.org/10.1126/science.1087143>.
- Brune, S., J. D. Keller, and S. Wahl, 2021: Evaluation of wind speed estimates in reanalyses for wind energy applications. *Adv. Sci. Res.*, **18**, 115–126, <https://doi.org/10.5194/asr-18-115-2021>.
- Büeler, D., R. Beerli, H. Wernli, and C. M. Grams, 2020: Stratospheric influence on ECMWF sub-seasonal forecast skill for energy-industry-relevant surface weather in European countries. *Quart. J. Roy. Meteor. Soc.*, **146**, 3675–3694, <https://doi.org/10.1002/qj.3866>.

- Buizza, R., and M. Leutbecher, 2015: The forecast skill horizon. *Quart. J. Roy. Meteor. Soc.*, **141**, 3366–3382, <https://doi.org/10.1002/qj.2619>.
- , M. Milleer, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, <https://doi.org/10.1002/qj.49712556006>.
- Cionni, I., J. Ramon, L. Lledó, H. Loukos, and T. Noël, 2018: Validation of observational dataset and recommendations to the energy users. Research and Innovation action H2020-SC5-2017 S2S4E Climate Services for Clean Energy Tech. Rep. 3.1, 106 pp., [https://s2s4e.eu/sites/default/files/2020-06/s2s4e\\_d31.pdf](https://s2s4e.eu/sites/default/files/2020-06/s2s4e_d31.pdf).
- Coelho, C. A., M. A. Firpo, and F. M. de Andrade, 2018: A verification framework for South American sub-seasonal precipitation predictions. *Meteor. Z.*, **27**, 503–520, <https://doi.org/10.1127/metz/2018/0898>.
- , B. Brown, L. Wilson, M. Mittermaier, and B. Casati, 2019: Forecast verification for S2S timescales. *Sub-Seasonal to Seasonal Prediction: The Gap between Weather and Climate Forecasting*, A. Robertson and F. Vitart, Eds., Elsevier, 337–361.
- Diro, G. T., and H. Lin, 2020: Subseasonal forecast skill of snow water equivalent and its link with temperature in selected SubX models. *Wea. Forecasting*, **35**, 273–284, <https://doi.org/10.1175/WAF-D-19-0074.1>.
- Dorrington, J., I. Finney, T. Palmer, and A. Weisheimer, 2020: Beyond skill scores: Exploring sub-seasonal forecast value through a case-study of French month-ahead energy prediction. *Quart. J. Roy. Meteor. Soc.*, **146**, 3623–3637, <https://doi.org/10.1002/qj.3863>.
- Fu, X., B. Wang, D. E. Waliser, and L. Tao, 2007: Impact of atmosphere–ocean coupling on the predictability of monsoon intraseasonal oscillations. *J. Atmos. Sci.*, **64**, 157–174, <https://doi.org/10.1175/JAS3830.1>.
- Hagedorn, R., R. Buizza, T. M. Hamill, M. Leutbecher, and T. N. Palmer, 2012: Comparing TIGGE multimodel forecasts with reforecast-calibrated ECMWF ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, **138**, 1814–1827, <https://doi.org/10.1002/qj.1895>.
- Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2).
- , and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hoskins, B., 2012: Predictability beyond the deterministic limit. World Meteorological Organization, <https://public.wmo.int/en/bulletin/predictability-beyond-deterministic-limit>.
- International Energy Agency, 2020: World energy outlook 2020. IEA Rep., 464 pp., <https://iea.blob.core.windows.net/assets/a72d8abf-de08-4385-8711-b8a062d6124a/WEO2020.pdf>.
- Jifan, C., 1989: Predictability of the atmosphere. *Adv. Atmos. Sci.*, **6**, 335–346, <https://doi.org/10.1007/BF02661539>.
- Jolliffe, I., and D. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons, 254 pp.
- Jones, C., D. E. Waliser, K. M. Lau, and W. Stern, 2004a: Global occurrences of extreme precipitation and the Madden–Julian Oscillation: Observations and predictability. *J. Climate*, **17**, 4575–4589, <https://doi.org/10.1175/3238.1>.
- , —, —, and —, 2004b: The Madden–Julian Oscillation and its impact on Northern Hemisphere weather predictability. *Mon. Wea. Rev.*, **132**, 1462–1471, [https://doi.org/10.1175/1520-0493\(2004\)132<1462:TMOAII>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1462:TMOAII>2.0.CO;2).
- Jourdi er, B., 2015: Wind resource in metropolitan France: Assessment methods, variability and trends. Ph.D. thesis, Ecole Polytechnique, 229 pp., [http://inis.iaea.org/search/search.aspx?orig\\_q=RN:48072202](http://inis.iaea.org/search/search.aspx?orig_q=RN:48072202).
- , 2020: Evaluation of ERA5, MERRA-2, COSMO-REA6, NEWA and AROME to simulate wind power production over France. *Adv. Sci. Res.*, **17**, 63–77, <https://doi.org/10.5194/asr-17-63-2020>.
- Jung, T., F. Vitart, L. Ferranti, and J.-J. Morcrette, 2011: Origin and predictability of the extreme negative NAO winter of 2009/10. *Geophys. Res. Lett.*, **10**, L16815, <https://doi.org/10.1029/2011GL046786>.
- Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.
- Koster, R. D., and Coauthors, 2011: The second phase of the global land–atmosphere coupling experiment: Soil moisture contributions to subseasonal forecast skill. *J. Hydrometeorol.*, **12**, 805–822, <https://doi.org/10.1175/2011JHM1365.1>.
- Leung, L. R., A. F. Hamlet, D. P. Lettenmaier, and A. Kumar, 1999: Simulations of the ENSO hydroclimate signals in the Pacific Northwest Columbia River Basin. *Bull. Amer. Meteor. Soc.*, **80**, 2313–2330, [https://doi.org/10.1175/1520-0477\(1999\)080<2313:SOTEHS>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<2313:SOTEHS>2.0.CO;2).
- Leutbecher, M., 2005: On ensemble prediction using singular vectors started from forecasts. ECMWF Tech. Memo. 462, 11 pp., <https://doi.org/10.21957/xuyeqtxv>.
- Lin, H., and Z. Wu, 2011: Contribution of the autumn Tibetan Plateau snow cover to seasonal prediction of North American winter temperature. *J. Climate*, **24**, 2801–2813, <https://doi.org/10.1175/2010JCLI3889.1>.
- Lledó, L., and F. J. Doblas-Reyes, 2020: Predicting daily mean wind speed in Europe weeks ahead from MJO status. *Mon. Wea. Rev.*, **148**, 3413–3426, <https://doi.org/10.1175/MWR-D-19-0328.1>.
- Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17**, 321–333, <https://doi.org/10.1111/j.2153-3490.1965.tb01424.x>.
- Lynch, K. J., D. J. Brayshaw, and A. Charlton-Perez, 2014: Verification of European subseasonal wind speed forecasts. *Mon. Wea. Rev.*, **142**, 2978–2990, <https://doi.org/10.1175/MWR-D-13-00341.1>.
- Machin, D., T. Bryant, D. Altman, and M. Gardner, 2013: *Statistics with Confidence: Confidence Intervals and Statistical Guidelines*. John Wiley and Sons, 254 pp.
- Manzanas, R., J. M. Gutiérrez, J. Bhend, S. Hemri, F. J. Doblas-Reyes, V. Torralba, E. Penabad, and A. Brookshaw, 2019: Bias adjustment and ensemble recalibration methods for seasonal forecasting: A comprehensive intercomparison using the C3S dataset. *Climate Dyn.*, **53**, 1287–1305, <https://doi.org/10.1007/s00382-019-04640-4>.
- Matheson, J. E., and R. L. Winkler, 1976: Scoring rules for continuous probability distributions. *Manage. Sci.*, **22**, 1087–1096, <https://doi.org/10.1287/mnsc.22.10.1087>.
- Molteni, F., U. Cubasch, and S. Tibaldi, 1986: 30- and 60-day forecast experiments with the ECMWF spectral models. *Workshop on Predictability in the Medium and Extended Range*, Shinfield Park, Reading, ECMWF, 51–107, <https://www.ecmwf.int/node/11210>.

- Monhart, S., C. Spirig, J. Bhend, K. Bogner, C. Schär, and M. A. Liniger, 2018: Skill of subseasonal forecasts in Europe: Effect of bias correction and downscaling using surface observations. *J. Geophys. Res. Atmos.*, **123**, 7999–8016, <https://doi.org/10.1029/2017JD027923>.
- Namias, J., 1952: The annual course of month-to-month persistence in climatic anomalies. *Bull. Amer. Meteor. Soc.*, **33**, 279–285, <https://doi.org/10.1175/1520-0477-33.7.279>.
- Palmer, T. N., 2012: Towards the probabilistic Earth-system simulator: A vision for the future of climate and weather prediction. *Quart. J. Roy. Meteor. Soc.*, **138**, 841–861, <https://doi.org/10.1002/qj.1923>.
- , R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo. 598, 42 pp., <https://doi.org/10.21957/ps8gbwbdv>.
- Ramon, J., L. Lledó, V. Torralba, A. Soret, and F. J. Doblas-Reyes, 2019: What global reanalysis best represents near-surface winds? *Quart. J. Roy. Meteor. Soc.*, **145**, 3236–3251, <https://doi.org/10.1002/qj.3616>.
- Raoult, B., C. Bergeron, A. L. Alós, J.-N. Thépaut, and D. Dee, 2017: Climate service develops user-friendly data store. *ECMWF Newsletter*, No. 151, ECMWF, Reading, United Kingdom, <https://www.ecmwf.int/en/newsletter/151/meteorology/climate-service-develops-user-friendly-data-store>.
- Robertson, A., and F. Vitart, 2018: *Sub-seasonal to Seasonal Prediction: The Gap between Weather and Climate Forecasting*. Elsevier, 512 pp.
- Sanders, F., 1963: On subjective probability forecasting. *J. Appl. Meteor.*, **2**, 191–201, [https://doi.org/10.1175/1520-0450\(1963\)002<0191:OSPF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1963)002<0191:OSPF>2.0.CO;2).
- Seneviratne, S., and Coauthors, 2012: Changes in climate extremes and their impacts on the natural physical environment. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. B. Field et al., Eds., Cambridge University Press, 109–230, [https://www.ipcc.ch/site/assets/uploads/2018/03/SREX-Chap3\\_FINAL-1.pdf](https://www.ipcc.ch/site/assets/uploads/2018/03/SREX-Chap3_FINAL-1.pdf).
- Simmons, A., and Coauthors, 2021: Low frequency variability and trends in surface air temperature and humidity from ERA5 and other datasets. ECMWF Tech. Memo. 881, 99 pp., <https://doi.org/10.21957/ly5vbtbfd>.
- Sobolowski, S., G. Gong, and M. Ting, 2010: Modeled climate state and dynamic responses to anomalous North American snow cover. *J. Climate*, **23**, 785–799, <https://doi.org/10.1175/2009JCLI3219.1>.
- Torralba, V., F. J. Doblas-Reyes, D. MacLeod, I. Christel, and M. Davis, 2017: Seasonal climate prediction: A new source of information for the management of wind energy resources. *J. Appl. Meteor. Climatol.*, **56**, 1231–1247, <https://doi.org/10.1175/JAMC-D-16-0204.1>.
- Unger, D. A., 1985: A method to estimate the continuous ranked probability score. *Conf. Prob. Stat. Atmos. Sci.*, **9**, 206–213, [https://jglobal.jst.go.jp/en/detail?JGLOBAL\\_ID=200902092398162270](https://jglobal.jst.go.jp/en/detail?JGLOBAL_ID=200902092398162270).
- van den Hurk, B., F. Doblas-Reyes, G. Balsamo, R. D. Koster, S. I. Seneviratne, and H. Camargo, 2012: Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe. *Climate Dyn.*, **38**, 349–362, <https://doi.org/10.1007/s00382-010-0956-2>.
- Vigaud, N., M. K. Tippett, J. Yuan, A. W. Robertson, and N. Acharya, 2019: Probabilistic skill of subseasonal surface temperature forecasts over North America. *Wea. Forecasting*, **34**, 1789–1806, <https://doi.org/10.1175/WAF-D-19-0117.1>.
- Vitart, F., 2014: Evolution of ECMWF sub-seasonal forecast skill scores. *Quart. J. Roy. Meteor. Soc.*, **140**, 1889–1899, <https://doi.org/10.1002/qj.2256>.
- , and Coauthors, 2017: The Subseasonal to Seasonal (S2S) prediction project database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173, <https://doi.org/10.1175/BAMS-D-16-0017.1>.
- , and Coauthors, 2019: Extended-range prediction. ECMWF Tech. Memo. 854, 60 pp., <https://doi.org/10.21957/pdivp3t9m>.
- White, C. J., and Coauthors, 2017: Potential applications of sub-seasonal-to-seasonal (S2S) predictions. *Meteor. Appl.*, **24**, 315–325, <https://doi.org/10.1002/met.1654>.
- Wilks, D. S., 2019: *Statistical Methods in the Atmospheric Sciences*. 4th ed. Elsevier, 840 pp.
- Woolnough, S. J., F. Vitart, and M. A. Balmaseda, 2007: The role of the ocean in the Madden–Julian Oscillation: Implications for MJO prediction. *Quart. J. Roy. Meteor. Soc.*, **133**, 117–128, <https://doi.org/10.1002/qj.4>.
- Žagar, N., and I. Szunyogh, 2020: Comments on “What is the predictability limit of midlatitude weather?”. *J. Atmos. Sci.*, **77**, 781–785, <https://doi.org/10.1175/JAS-D-19-0166.1>.
- Zhang, F., Y. Q. Sun, L. Magnusson, R. Buizza, S.-J. Lin, J.-H. Chen, and K. Emanuel, 2019: What is the predictability limit of midlatitude weather? *J. Atmos. Sci.*, **76**, 1077–1091, <https://doi.org/10.1175/JAS-D-18-0269.1>.
- Zhu, H., M. C. Wheeler, A. H. Sobel, and D. Hudson, 2014: Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Mon. Wea. Rev.*, **142**, 1556–1569, <https://doi.org/10.1175/MWR-D-13-00222.1>.

# 3 Improving sub-seasonal predictions

Continuous improvement is better than  
delayed perfection.  
— Mark Twain

## Objective

The main objective of this chapter is to improve sub-seasonal predictions of 100-m wind speed and 2-m temperature over Europe by statistically downscaling information from large-scale, upper-level fields.

## Data and Methods

In addition to the data employed in the previous chapter, we use geopotential height at 500 hPa (Z500) over Euro-Atlantic from the European Centre for Medium-Range Weather Forecasts (ECMWF). We first employ Redundancy Analysis to obtain spatial patterns of variability of Z500 conditioned on the surface fields. We then apply the relationship between Z500 patterns and the surface fields on sub-seasonal predictions of Z500 to obtain statistical predictions of surface fields. Subsequently, we combine statistical predictions of surface fields with their dynamical counterparts to obtain hybrid predictions.

## Key conclusions

- The large-scale patterns obtained using Redundancy Analysis have a higher explanatory power than those obtained using Principal Component Analysis for predicting surface fields over Europe;
- The combination of dynamical and statistical predictions significantly improves the skill horizon of sub-seasonal surface field predictions over a large part of Europe;

## Publication

This chapter, which was submitted to *Monthly Weather Review* in June 2022, has been accepted for publication in October 2022 with an assigned DOI (<https://doi.org/10.1175/MWR-D-22-0170.1>) (©**American Meteorological Society**).

## Statistical downscaling to improve the sub-seasonal predictions of energy-relevant surface variables

NAVEEN GOUTHAM,<sup>a,b</sup> RIWAL PLOUGONVEN,<sup>b</sup> HIBA OMRANI,<sup>a</sup> ALEXIS TANTET,<sup>b</sup> SYLVIE PAREY,<sup>a</sup> PETER TANKOV,<sup>c</sup>  
PETER HITCHCOCK,<sup>d</sup> AND PHILIPPE DROBINSKI<sup>b</sup>

<sup>a</sup> EDF Lab Paris-Saclay, Palaiseau, France.

<sup>b</sup> Laboratoire de Météorologie Dynamique-IPSL, Ecole Polytechnique, Institut Polytechnique de Paris, ENS, PSL Research University, Sorbonne Université, CNRS, France.

<sup>c</sup> CREST/ENSAE, Institut Polytechnique de Paris, Palaiseau, France.

<sup>d</sup> Dept. Earth and Atmospheric Sciences, Cornell University, Ithaca, New York, USA.

**ABSTRACT:** Owing to the increasing share of variable renewable energies in the electricity mix, the European energy sector is increasingly becoming weather sensitive. In this regard, skillful sub-seasonal predictions of essential climate variables can provide considerable socio-economic benefits to the energy sector. The aim of this study is therefore to improve the European sub-seasonal predictions of 100 m wind speed and 2 m temperature, which we achieve through statistical downscaling. We employ Redundancy Analysis (RDA) to estimate spatial patterns of variability from large-scale fields that allow for the best prediction of surface fields. We compare explanatory powers between the patterns obtained using RDA against those derived using Principal Component Analysis (PCA), when used as predictors in multi-linear regression models to predict surface fields, and show that the explanatory power of the former is superior to that of the latter. Subsequently, we employ the estimated relationship between RDA patterns and surface fields to produce statistical probabilistic predictions of gridded surface fields using dynamical ensemble predictions of RDA patterns. We finally demonstrate how a simple combination of dynamical and statistical predictions of surface fields significantly improves the accuracy of sub-seasonal predictions of both variables over a large part of Europe. We attribute the improved accuracy of these combined predictions to improvements in reliability and resolution.

### 1. Introduction

Sub-seasonal predictions refer to predictions beyond two weeks and up to two months (Robertson and Vitart 2018). These predictions are influenced by both atmospheric initial conditions and boundary forcings (Hoskins 2012). Predictability on sub-seasonal timescales is limited by the use of imperfect initial conditions and imperfect numerical formulations in prediction models (Lorenz 1963, 1982; Palmer et al. 2009; Leutbecher et al. 2016). Predictability of fine-scale atmospheric features on sub-seasonal timescales remains poor for fundamental reasons, because of the chaos inherent in the atmosphere (Lorenz 1965; Jifan 1989; Zhang et al. 2019a). However, the predictability of large-scale, low-frequency features in the ocean, over land, and in the cryosphere lasts well beyond two weeks (Vitart et al. 2012; Buizza and Leutbecher 2015; Toth and Buizza 2019). The key sources of sub-seasonal predictability are Madden-Julian Oscillation (e.g., Jones et al. 2004a,b; Zheng et al. 2018), snow cover (e.g., Sobolowski et al. 2010; Lin and Wu 2011; Orsolini et al. 2013), stratosphere-troposphere interaction (e.g., Baldwin et al. 2003; Domeisen et al. 2020; Schwartz and Garfinkel 2020), land conditions (e.g., Koster et al. 2011; van den Hurk et al. 2012; Prodhomme et al. 2016; Seo et al. 2019), and ocean conditions (e.g., Woolnough et al. 2007; Fu

et al. 2007; Subramanian et al. 2019). Predictions on sub-seasonal timescales however need to be averaged on large enough spatiotemporal scales to extract relevant and predictable components of the signal (Lorenz 1982; Zhu et al. 2014; Buizza and Leutbecher 2015). Since sub-seasonal predictions are beyond deterministic limits of predictability (i.e., about ten days), these predictions are produced as ensembles of numerical integrations, describing a range of possibilities instead of a unique best estimate of the future state. This shift from determinism to probabilism has been a major breakthrough in extending the predictability horizon of sub-seasonal predictions (Palmer 2012).

With a transition towards low carbon energy systems, the energy industry is going to be one of the most important end-users of sub-seasonal predictions (White et al. 2017). Skillful sub-seasonal predictions of essential climate variables such as wind speed, solar radiation, and surface temperature can inform the energy industry about expected renewable energy production and energy consumption, and further prepare the sector for any possible risks which may arise due to anomalies. A non-exhaustive list of applications in the energy sector for which sub-seasonal predictions can be instrumental includes determining required reserve levels, maintenance scheduling, assessment of extreme risks, determining grid transmission capacity, and trading electricity in power markets.

Europe, being one of the world's largest energy consuming and greenhouse gas emitting regions, sits at the fore-

---

Corresponding author: Naveen Goutham,  
naveen.goutham@edf.fr/naveen.goutham@polytechnique.edu

front of energy transition (Liobikienė and Butkus 2017; Jonek-Kowalska 2022). In the European Union, wind power is becoming the largest renewable source of electricity (IEA 2020). Hence, we focus this study on sub-seasonal predictions of 100 m wind speed and 2 m temperature over Europe. Several studies have assessed the quality of sub-seasonal predictions of wind speed and surface temperature over Europe, and have found skillful predictions relative to climatology for weekly mean quantities beyond two weeks (e.g., Lynch et al. 2014; Monhart et al. 2018; Diro and Lin 2020; Dorrington et al. 2020; Goutham et al. 2022). Although the fundamental sources of sub-seasonal predictability have been identified (Vitart et al. 2012), the physical relationships between large-scale, low-frequency fields and surface fields (i.e., within the planetary boundary layer) are not well represented in sub-seasonal prediction models due to parametrizations (Palmer et al. 2009; Leutbecher et al. 2016; Robertson and Vitart 2018; Lledó and Doblas-Reyes 2020). In addition, the forecast errors of surface fields grow relatively faster than that of large-scale fields due to increased sensitivity of the former to model parametrizations (e.g., Buizza and Leutbecher 2015; Toth and Buizza 2019). Given the longer skill horizon of large-scale fields compared to surface fields (Buizza et al. 2015; Toth and Buizza 2019; Büeler et al. 2021), there is an opportunity to improve sub-seasonal surface-field predictions by accounting for the misrepresentations in physical relationships between large-scale and surface-fields using historical data. In other words, the information contained in the prediction of large-scale fields is more reliable than that in surface fields, and statistical downscaling techniques can be implemented to correctly transfer this information from large-scale fields to surface fields (e.g., Manzanar et al. 2018; Goutham et al. 2021).

The most popular statistical downscaling techniques are the linear methods due to their transparency and ease of interpretation (Benestad et al. 2008; Wilks 2019). Generally, linear statistical downscaling is done in three stages; one, choosing predictors which have physical relationships with the predictand; two, obtaining the linear relationship between predictors and the predictand; and finally, using future dynamical predictions of predictors to reconstruct the predictand. In a majority of studies focusing on statistical downscaling of surface variables over Europe, weather regimes obtained from dimension reduction or clustering of geopotential height at 500 hPa (Z500) are used as predictors, which are then regressed on surface variables (e.g. Grams et al. 2017; Alonzo et al. 2017; Ramon et al. 2021). The obtained coefficients are then employed on future predictions of weather regimes to reconstruct surface fields. Z500 has long been the variable of choice to determine weather regimes as it represents the mid-troposphere, making it easier to capture large-scale flow (Wallace and Gutzler 1981; Cheng and Wallace 1993; Wilby and Wigley 1997; Plaut and Simonnet 2001;

Alonzo et al. 2017). Alonzo et al. (2017) have developed a methodology to estimate the distribution of surface wind speed over France based on the knowledge (or forecast) of the large-scale atmospheric state, the latter being summarized by the first few patterns obtained through Principal Component Analysis (PCA). It was verified that these patterns or Empirical Orthogonal Functions (EOFs) represent classical Euro-Atlantic weather regimes. Although each weather regime is associated with a set of surface meteorological conditions (van der Wiel et al. 2019), the main limitations of the use of classical weather regimes for predicting surface fields are that these weather regimes represent large-scale atmospheric variability independently of the predictand and that each surface climate variable responds differently to the same weather regime (Bloomfield et al. 2019). This calls for the development of new approaches to obtain large-scale spatial patterns of variability that take into account variability of the predictand itself (Bloomfield et al. 2019). One such approach is presented in Bloomfield et al. (2019) where they use k-means clustering to find "targeted circulation types" conditioned on the European power system. It is important to understand large-scale flow patterns that have the highest impact on surface variables as these patterns can be used to enhance the skill horizon of specific surface variables. The objectives of this research are therefore to identify spatial patterns of variability of Z500 conditioned on 100 m wind speed and 2 m temperature over Europe, and to use ensemble predictions of these patterns to improve sub-seasonal ensemble predictions of 100 m wind speed and 2 m temperature.

In this study, we employ a multivariate statistical technique called Redundancy Analysis (RDA) between the Z500 field and the surface fields to obtain patterns of Z500 that maximize explained variance of the surface variables (von Storch et al. 1999; Tippett et al. 2008; Wilks 2014, 2019). We then apply the estimated linear regression coefficients on ensemble dynamical predictions of a restricted number of RDA patterns of Z500 to obtain statistical probabilistic predictions of surface variables. Several dimension reduction methods exist to summarise coupled variations of large-scale fields and surface fields using a few patterns and their corresponding coefficients (von Storch et al. 1999; Tippett et al. 2008; Wilks 2019). Among these methods, Redundancy Analysis distinguishes itself from classical multivariate techniques such as Canonical Correlation Analysis or Maximum Covariance Analysis by being asymmetric in the treatment of predictor and predictand (i.e., it distinguishes dependent and independent variables), as is the case with multi-linear regression (von Storch et al. 1999; Wang and Zwiers 2001; Tippett et al. 2008; Wilks 2014). As far as the authors are aware, this is the first study to compare explanatory power between the patterns obtained using RDA of Euro-Atlantic Z500 against those derived using PCA, when used as predictors in a multi-linear regression model to predict 100 m wind



speed and 2 m temperature over Europe, and also the first to reveal the RDA patterns of Z500 conditioned on the same two surface climate variables. In addition, contrary to several studies which have discarded dynamical predictions of surface variables completely in favor of statistical predictions (e.g., Alonzo et al. 2017; Ramon et al. 2021), we demonstrate how a simple combination of dynamical predictions of surface variables with statistical predictions derived from redundancy analysis can enhance prediction skill. Although the idea of combining dynamical and statistical predictions has already been illustrated in some recent studies on seasonal timescales (e.g., Schepen et al. 2012, 2014, 2016; Strazzo et al. 2019), in this study, we demonstrate the value gained through a combination on sub-seasonal timescales. We also explore forecast quality attributes of different ensemble predictions to identify those that lead to differences in predictive quality between dynamical and combined (i.e., dynamical + statistical) predictions.

The article is organised as follows: Section 2 outlines the data used; Section 3 describes Redundancy Analysis, the combination of dynamical and statistical predictions, and the metrics used to evaluate quality of predictions; Section 4 presents the results in three parts: (a) compares and contrasts patterns obtained using RDA against those of PCA, (b) compares the quality of different ensemble predictions, and (c) takes a closer look at forecast attributes that contribute to differences in prediction quality between different ensemble predictions; Sections 5 and 6 are reserved for discussions and conclusions, respectively.

## 2. Data

### a. Forecasts and Re-forecasts

The forecasts and retrospective forecasts (re-forecasts) data used in this study originate from extended-range predictions (Vitart et al. 2017) of the European Centre for Medium-Range Weather Forecasts (ECMWF). The medium-range (i.e., up to two weeks) ocean-atmosphere coupled ensemble forecasts are extended to 46 days twice a week at 00 UTC on Mondays and Thursdays to produce extended-range ensemble predictions (Vitart et al. 2019). The operational ensemble predictions consist of 51 members (50 perturbed + control). The perturbed members are obtained using singular vectors (Leutbecher 2005; Leutbecher and Palmer 2008) and ensemble data assimilation (Buizza et al. 2008; Isaksen et al. 2010). Stochastically Perturbed Parametrisation Tendencies (SPPT) scheme is used to represent model uncertainty (Buizza et al. 1999; Palmer et al. 2009; Leutbecher et al. 2016). These predictions are originally issued at a spatial resolution of Tco639L91 (~18 km) up to a lead time of 15 days, and at Tco319L91 (~36 km) after (Vitart et al. 2017, 2019).

The operational prediction model begins to drift significantly from reality after about ten days of coupled integrations. This drift can be attributed to inherent atmospheric unpredictability (Zhang et al. 2019b; Žagar and Szunyogh 2020), and the use of imperfect initial conditions and imperfect representation of physical processes in the numerical model (Palmer et al. 2009; Leutbecher et al. 2016). It is imperative to remove the drift before employing the model. The ECMWF produces re-forecasts to estimate and remove the operational model drift (Vitart et al. 2008). A re-forecast set consists of ensemble forecasts of 11 members (10 perturbed + control) issued for the same calendar day of the year as the operational forecast over each of the past 20 years. ERA5 reanalysis provides the initial conditions for the re-forecasts. This re-forecast set with 220 integrations (20 years  $\times$  11 members) allows for evaluation of the model climatology of operational forecasts.

We retrieve forecasts and the corresponding re-forecasts of 2 m temperature (T2m), zonal and meridional components of 100 m wind speed, and geopotential at 500 hPa issued during boreal winter (DJF) on a global grid between December 2016 and February 2020. The retrieved spatial resolution is  $0.9^\circ$  and the temporal resolution is 6 h (instantaneous values at 00, 06, 12, and 18 UTC). The data are retrieved from the Meteorological Archival and Retrieval System (MARS) of the ECMWF. The 100 m wind speed (U100) is computed as the square root of the sum of squares of zonal and meridional components. The geopotential height (Z500) is computed by dividing the geopotential by the Earth's gravitational acceleration,  $g$  ( $= 9.806 \text{ m s}^{-2}$ ). As the prediction model is undergoing periodic improvements, the dataset used in this study consists of forecasts and re-forecasts from several versions (CY43R1, CY43R3, CY45R1, and CY46R1) (Vitart et al. 2019). Nevertheless, the differences in model formulation and hence the statistics between different versions are marginal (refer to Appendix A in Goutham et al. (2022)). We focus on boreal winter in this study as this season experiences high variability in wind energy production in addition to increased energy demand mainly for space heating. Furthermore, predictions are more skillful in winter compared with other seasons due to stronger boundary conditions (e.g., sea surface temperature gradients), reinforced coupling (e.g., stratosphere-troposphere), and enhanced memory of initial conditions such as soil moisture among others (Robertson and Vitart 2018). In this study, only the perturbed members of forecasts and re-forecasts are used. The reader is referred to the data availability statement to learn about the missing control member. All the results shown in this study involving operational predictions rely on re-forecasts for calibration as explained in Appendix a.

### b. Reference

Generally, the forecast quality is assessed by comparing against observations (Coelho et al. 2019; Wilks 2019). However, in the absence of a serially complete and spatially coherent observed data set, reanalysis is used as a reference in forecast verification (Kalnay 2003). In this study, we use ERA5 reanalysis (Hersbach et al. 2020) as reference. ERA5 reanalysis is a fifth generation high resolution (hourly output, 31 km horizontal grid spacing) reanalysis produced using 4D-Var data assimilation and the CY41R2 version of the Integrated Forecast System of the ECMWF (Hersbach et al. 2020). We retrieve ERA5 reanalysis of T2m, zonal and meridional components of 100 m wind speed, and geopotential at 500 hPa on a global grid between January 1979 and January 2021 at the same spatial and temporal resolution as the forecasts. The data are retrieved from the Climate Data Store of the Copernicus Climate Change Services (Raoult et al. 2017). The 100 m wind speed and geopotential height are computed as previously described. Although ERA5 reanalysis shows cold biases in representing surface temperature over the Iberian peninsula and the Mediterranean (Johannsen et al. 2019), it represents the means and extremes well over most of Europe (e.g., Simmons et al. 2021; Velikou et al. 2022). Although ERA5 reanalysis severely underestimates the mean winds over complex terrain, it represents the variability of wind speed more realistically compared with other reanalysis datasets over Europe (e.g., Ramon et al. 2019; Jourdiier 2020; Dörenkämper et al. 2020; Brune et al. 2021; Molina et al. 2021; Murcia et al. 2022). In spite of the biases, the representation errors of ERA5 reanalysis are small, and hence acceptable for verification (Ramon et al. 2019; Velikou et al. 2022) and statistical modeling (Tarek et al. 2019). Accordingly, ERA5 reanalysis is used as a reference in forecast verification and as well as for training the statistical model in this study.

## 3. Methodology

### a. Redundancy Analysis

Redundancy Analysis (RDA) is a multivariate statistical technique that attempts to find lower-dimensional patterns of linear dependence between two multivariate data sets (i.e., between predictor and predictand) maximizing the coefficient of determination of linear regression (von Storch et al. 1999; Wang and Zwiers 2001; Tippett et al. 2008; Wilks 2014). There exist several other methods to find linearly coupled patterns between two multivariate datasets, notably Canonical Correlation Analysis (CCA) and Maximum Covariance Analysis (MCA). RDA, unlike CCA or MCA, is asymmetric in the treatment of two data sets as it identifies one as the predictor and the other as the predictand. This way, the patterns derived from RDA

are specifically tailored for use in multi-linear regression models.

$$\mathbf{P} = \begin{pmatrix} p_{1,1} & p_{1,2} & \dots & p_{1,t} \\ p_{2,1} & \cdot & \dots & p_{2,t} \\ \cdot & \cdot & \dots & \cdot \\ p_{m,1} & \cdot & \dots & p_{m,t} \end{pmatrix}$$

and

$$\mathbf{Q} = \begin{pmatrix} q_{1,1} & q_{1,2} & \dots & q_{1,t} \\ q_{2,1} & \cdot & \dots & q_{2,t} \\ \cdot & \cdot & \dots & \cdot \\ q_{n,1} & \cdot & \dots & q_{n,t} \end{pmatrix}$$

Let  $\mathbf{P}$  be the predictor anomaly matrix with each column representing an observation at each of the  $m$  grid points. Let  $\mathbf{Q}$  be the predictand anomaly matrix with each column representing an observation at each of the  $n$  grid points. For illustration purpose,  $\mathbf{P}$  can be thought of as Z500, and  $\mathbf{Q}$  as U100. The elements of both  $\mathbf{P}$  and  $\mathbf{Q}$  are weighted by square root of cosine of latitude to equalize variance (von Storch et al. 1999; Wilks 2014, 2019). We use gridded Z500 weekly mean anomalies over Euro-Atlantic ( $20^\circ$  N -  $80^\circ$  N,  $120^\circ$  W -  $40^\circ$  E) as the predictor and gridded T2m/U100 weekly mean anomalies over Europe ( $34^\circ$  N -  $74^\circ$  N,  $13^\circ$  W -  $40^\circ$  E) as predictand in this study. The choice of the predictor domain and its sensitivity to the predictand domain is discussed in section 5a. Regarding the calculation of anomalies, we tested lagging 15-year as well as 20-year mean climatology for computing anomalies. We observed that using lagging 15-year climatology (i.e., the most recent 15-year period as climatology) performs better relative to using lagging 20-year climatology (i.e., the most recent 20-year period) in alleviating cold biases in temperature forecasts that are derived from the ongoing climate warming (not shown). This observation is consistent with the ones previously seen in the literature (e.g., Wilks 2013; Wilks and Livezey 2013; Wilks 2014). Therefore, we compute Z500, T2m, and U100 anomalies by removing lagging 15-year mean climatology from the observed weekly mean. The climatological data used to compute anomalies correspond to the same week and month of the year as the observation. Although U100 shows no particular trend, we retain the 15-year period for computing U100 anomalies to make the inter-variable comparison consistent. We have more explanatory variables (i.e., grid points) than the number of observations in matrices  $\mathbf{P}$  and  $\mathbf{Q}$ . Hence, to prevent over-determination and to lessen the computational burden, we perform Principal Component Analysis (PCA) of matrices  $\mathbf{P}$  and  $\mathbf{Q}$  to obtain their corresponding principal components (PC) retaining 99% of the variance in the original data (von Storch et al. 1999; Wilks 2019). As the predictor and the predictand vectors are measured in different units, we normalize them by subtracting the grid-point mean and dividing by



the grid-point standard deviation (Wilks 2019). The predictor and predictand PCs of the normalized variables  $\mathbf{P}'$  and  $\mathbf{Q}'$  are computed as  $\mathbf{X} = \mathbf{E}_P^T \mathbf{P}'$  and  $\mathbf{Y} = \mathbf{E}_Q^T \mathbf{Q}'$ , respectively. Here, the matrices  $\mathbf{E}_P^T$  and  $\mathbf{E}_Q^T$  hold the predictor and predictand patterns, respectively. The superscript T denotes vector or matrix transpose. Using centered variables in place of normalized variables marginally degrades the results (not shown). The joint sample variance-covariance matrix of the leading predictor and predictand PCs is given by

$$\begin{aligned} \mathbf{S} &= (\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_i; \mathbf{y}_1; \mathbf{y}_2; \dots; \mathbf{y}_j) (\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_i; \mathbf{y}_1; \mathbf{y}_2; \dots; \mathbf{y}_j)^T \\ &= \begin{pmatrix} \mathbf{S}_{XX} & \mathbf{S}_{XY} \\ \mathbf{S}_{YX} & \mathbf{S}_{YY} \end{pmatrix}. \end{aligned} \quad (1)$$

Here, the  $(i + j)$  by  $t$  matrix  $(\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_i \mathbf{y}_1 \mathbf{y}_2 \dots \mathbf{y}_j)^T$  is formed through concatenation of the leading  $i$  predictor PCs  $\mathbf{x}_i$  and the leading  $j$  predictand PCs  $\mathbf{y}_j$ . Since we use standardized variables,  $\mathbf{S}$  is in fact a correlation matrix (von Storch et al. 1999; Wilks 2019). Nonetheless, we use "covariance matrix" as a general terminology to describe the method. The covariance matrix of predictand PCs conditioned on predictor PCs is given by

$$\mathbf{S}_{Y|X} = \mathbf{S}_{YX} \mathbf{S}_{XX}^{-1} \mathbf{S}_{XY}. \quad (2)$$

The eigen-decomposition of the square symmetric matrix in equation 2 yields orthonormal eigenvectors  $\mathbf{B}$  and diagonal matrix  $\Lambda$  of positive eigenvalues  $\lambda$ , both sorted in descending order based on the values of  $\lambda$ . The columns of  $\mathbf{B}$  consist in the patterns that account for the variance of predictand PCs when conditioned on predictor PCs. We can deduce the predictor patterns  $\mathbf{A}$  through the equation

$$\mathbf{A} = \frac{1}{\sqrt{\Lambda}} \mathbf{S}_{XX}^{-1} \mathbf{S}_{XY} \mathbf{B}. \quad (3)$$

We can conveniently compute the predictor and predictand redundancy PCs using  $\mathbf{V} = \mathbf{A}^T \mathbf{X}$  and  $\mathbf{W} = \mathbf{B}^T \mathbf{Y}$ , respectively. The regression coefficients of the linear relationship between  $\mathbf{V}$  and  $\mathbf{W}$  are given by  $\mathbf{R} = \sqrt{\Lambda}$ . The redundancy PCs  $\mathbf{V}$  and  $\mathbf{W}$  are linked through  $\mathbf{W} = \mathbf{R} \mathbf{V}$ . For a given number of retained patterns, redundancy analysis guarantees that the coefficient of determination of the linear regression is maximized. In this study, we use 7-day rolling averages of ERA5 reanalysis of Z500, T2m, and U100 in a Perfect Prognosis framework for fitting the model (e.g., Hewitson and Crane 1996; Zorita and von Storch 1999; Ramon et al. 2021). We compute regression coefficients by fitting a separate model for each predictand. We choose the training period to be the boreal winter between December 1999 and February 2016 (i.e., 17 years).

### b. Statistical and Hybrid Predictions

We can use redundancy regression coefficients ( $\mathbf{R}$ ), predictor patterns ( $\mathbf{A}$ ), predictand patterns ( $\mathbf{B}$ ), and the relationship between predictor and predictand redundancy PCs to predict the predictand given a new set of predictors. Let  $\mathbf{X}_o$  be a new set of predictor PCs computed from a new set of predictor anomaly matrix  $\mathbf{P}_o$  (i.e.,  $\mathbf{X}_o = \mathbf{E}_P^T \mathbf{P}'_o$  with  $\mathbf{E}_P^T$  unchanged from the initial analysis). We can compute the predictand redundancy PCs as  $\hat{\mathbf{W}} = \mathbf{R}^T \mathbf{V}_o = \mathbf{R}^T \mathbf{A}^T \mathbf{X}_o$ . We can then obtain the standardized predictand vector anomalies using the equation

$$\hat{\mathbf{Q}}_o' = \mathbf{E}_Q (\mathbf{B}^T)^{-1} \hat{\mathbf{W}} = \mathbf{E}_Q (\mathbf{B}^T)^{-1} \mathbf{R}^T \mathbf{A}^T \mathbf{X}_o. \quad (4)$$

We consider an ensemble of weekly mean anomalies of Z500 operational extended-range predictions from ECMWF at any given lead time as  $\mathbf{P}_o$ . The operational predictions are bias-corrected using the Mean and variance Adjustment method (Torrallba et al. 2017; Manzanos et al. 2019; Goutham et al. 2022) as described in Appendix a. The prediction anomalies are computed in a similar way to the observed anomalies, but using a lagging 15-year climatology derived from the re-forecasts. We apply equation 4 on a *restricted* number of PCs ( $\mathbf{X}_o$ ) of  $\mathbf{P}_o$  to obtain an ensemble of predicted weekly mean anomalies of T2m or U100. The truncation of predictor PCs is a necessary step to optimize the accuracy of ensemble predictions, and it will be discussed further in the following sections. The predicted T2m or U100 anomalies are converted to absolute values by adding the lagging 15-year climatology of the respective variable derived from re-forecasts. Although we use dynamical predictions of Z500 to predict surface fields, we refer to the predicted vectors as statistical (ST) predictions, emphasizing the role of the statistical relationship between the predictor and the predictand. We then obtain a 100-member ensemble hybrid (HY) prediction by concatenating a 50-member statistical surface field prediction with a 50-member dynamical (DY) surface field prediction. All the ensemble members of the hybrid prediction receive equal weights.

### c. Measures of prediction skill

Evaluation of probabilistic prediction skill involves measuring different aspects of prediction quality (Jolliffe and Stephenson 2003; Wilks 2019; Coelho et al. 2019). The most important attributes of a forecast/prediction quality are:

- Accuracy: it measures the average distance between forecasts and observations;
- Association: it measures the strength of the relationship between forecasts and observations;
- Reliability: it measures calibration of the issued forecast probabilities;

- Resolution: it measures how the frequency of occurrence of an event varies as the issued forecast probability changes; and
- Sharpness: it measures the ability of forecasts to produce concentrated predictive distributions that are distinct from climatological probabilities.

Several scores have been proposed in the literature to assess probabilistic prediction skill taking into account these different forecast attributes (e.g., Jolliffe and Stephenson 2003). In this study, we employ the following metrics and diagnostic plots:

#### 1. Continuous Ranked Probability Skill Score (CRPSS):

The Continuous Ranked Probability Score (CRPS) measures the distance between the Cumulative Distribution Functions (CDF) of a probabilistic prediction and an observation (Matheson and Winkler 1976; Unger 1985; Hersbach 2000). The CRPS is a negatively oriented score in that the smallest values indicate more accurate predictions. It is also a proper score as it rewards those predictions whose probabilities are concentrated around the observation (Gneiting and Raftery 2007). The CRPS has the same units as the physical quantity being verified. The CRPS can be decomposed into components consisting of reliability, resolution, and uncertainty (Hersbach 2000). The CRPSS compares the prediction skill of a given prediction system with that of a benchmark. In the absence of reliable forecasts for end-user applications, a common practice in the energy industry is to use observed climatology, a long-term average of observed weather (typically 35 years), as the expected weather. In this study, we use a 35-year lagging observed climatology, derived from ERA5 reanalysis, as a 35-member ensemble benchmark prediction (CL). This climatological data corresponds to the same week and month of the year as the dynamical prediction but taken over the last 35 years. The choice of a 15-year lagging climatology for computing anomalies, as described in section 3a, is solely to alleviate cold bias in statistical T2m predictions. Since the prediction systems compared in this study (i.e., dynamical, statistical, hybrid, and climatological) are composed of different ensemble sizes, we compute Fair-CRPS (FCRPS) and Fair-CRPSS (FCRPSS) to have an unbiased estimate of the scores (Ferro 2014). Skillful predictions should have FCRPSS greater than zero. The standard practice in forecast verification is to compute scores for re-forecasts, and use these scores as an indication of the skill of the operational forecasts (Jolliffe and Stephenson 2003). However, Goutham et al. (2022) have shown that the skill of operational predictions on S2S timescales over Europe is higher

than that of re-forecasts. The improved skill of operational predictions is mainly attributed to their larger ensemble size relative to the re-forecasts. Therefore, we compute all the scores and diagnostic plots for operational dynamical predictions and the corresponding statistical and hybrid predictions in this study. In particular, we first calculate the FCRPS of weekly averaged dynamical, statistical, hybrid, and climatological predictions for each of the forecast issue dates and at each of the considered lead times. We then compute FCRPSS and its mean over all the forecast issue dates using climatological predictions as the benchmark. We apply Wilcoxon signed-rank test (Wilcoxon 1945; Conover 1971; Wilks 2019) (Appendix b) to investigate the statistical significance of the differences of FCRPSS between hybrid and dynamical predictions.

#### 2. Proportion of Skillful Forecasts (PSF):

As the mean of a distribution is sensitive to the existence of outliers, the mean-FCRPSS overemphasizes negative instances, and can therefore lead to underestimation of the prediction skill (Goutham et al. 2022). Therefore, we compute Fair-Proportion of Skillful Forecasts (FPSF) in addition to mean-FCRPSS. As the name suggests, the FPSF is a proportion of the number of predictions that have FCRPSS greater than zero to the total number of predictions considered.

#### 3. Anomaly Correlation Coefficient (ACC):

The ACC is a deterministic score that measures the linear association as Pearson's correlation coefficient between the anomalies of the ensemble mean predictions and observations (Namias 1952; Wilks 2019). ACC is complementary to CRPS as it is insensitive to forecast errors in accuracy. Accordingly, a forecast can be skillful (based on ACC) if it has some temporal association with the observations, irrespective of the magnitude of its accuracy.

#### 4. Reliability diagram:

A reliability diagram is a diagnostic plot to understand the full joint distribution of predictions and observations for probabilistic predictions of a binary predictand (Sanders 1963; Jolliffe and Stephenson 2003; Wilks 2019). It can be used to measure reliability, resolution, and sharpness. In this study, we plot reliability diagrams for upper and lower terciles of weekly mean predictions averaged over a geographical domain. Geographical domain averaging, wherever applicable, is computed as the mean of cosine-latitude weighted grid-point values.

In this study, we compute all the scores and diagnostic plots for weekly averaged quantities at four sub-seasonal lead times. More specifically, lead week-3 corresponds

to the weekly average between days 14-20, week-4 between days 21-27, week-5 between days 28-34, and week-6 between days 35-41. We employ leave-one-out cross-validation to estimate the optimum number of predictor patterns (i.e., truncation) in statistical predictions. The truncation is carried out with the criterion to optimize the median of FCRPSS of target predictions (i.e., statistical or hybrid) over the domain. This means that the number of predictor patterns retained in the best statistical predictions and in statistical predictions which form a component of hybrid predictions is different. Since hybrid predictions retain a large part of the information from dynamical predictions, the statistical predictions which form a component of hybrid predictions require deep truncation, i.e. only a small number of predictor patterns are sufficient. On the other hand, a large number of predictor patterns are required to obtain optimum statistical-only predictions. We prefer, as a truncation criterion, the median of FCRPSS to the mean as the mean is sensitive to extreme values. The number of predictor principal components required to obtain optimum U100 and T2m predictions, and the sensitivity of these predictions to the number of retained principal components will be discussed in the following section.

#### 4. Results

As a first step, we test the efficiency of patterns obtained using PCA on one hand, to those obtained using RDA on the other, to provide information on surface fields when used as predictors in a multi-linear regression model. Subsequently, we analyze the differences in prediction quality between dynamical, statistical, hybrid, and climatological predictions as well as between U100 and T2m. Finally, we explore and compare several forecast quality attributes between hybrid and dynamical predictions to understand the reasons for the differences in accuracy between the two.

##### *a. How do the patterns derived using Redundancy Analysis differ from those obtained using Principal Component Analysis?*

We compare the differences between Empirical Orthogonal Functions (EOFs) of Z500 anomalies over Euro-Atlantic derived using Principal Component Analysis (PCA), against those derived using Redundancy Analysis (RDA) conditioned on U100 and T2m over Europe in Figure 1. The EOFs of the Z500 field in PCA are chosen independently of the predictand to represent the maximum possible variability contained in the predictor itself. In contrast, the EOFs of the Z500 field in RDA are chosen to maximize the explained variance of the predictand. Alonzo et al. (2017) have verified that the principal components obtained through PCA of Z500 over the Euro-Atlantic represent the classical Euro-Atlantic weather regimes.

The first three patterns shown in Figure 1a resemble classical weather regimes of North Atlantic Oscillation, Scandinavian regime, and Atlantic regime, respectively (Alonzo et al. 2017; Bloomfield et al. 2019; van der Wiel et al. 2019; Garrido-Perez et al. 2020). Please note that the patterns in Figure 1 are sign indefinite, and that the color bars have no units as the units are carried by the PCs. The imprints of weather regimes on 10 m wind speed and T2m can be obtained from Bloomfield et al. (2019), van der Wiel et al. (2019), and Garrido-Perez et al. (2020). The first observation that can be made from Z500 patterns in Figure 1 is that the centers of action of RDA (top rows in (b) and (c)) are shifted towards or onto the European domain. This is logical as RDA patterns are conditioned on U100 and T2m over the European domain. Besides, we can notice variations in the strengths of troughs and ridges between PCA and RDA Z500 patterns. The Z500 patterns obtained using RDA further display inter-variable differences which may be attributed to the behavior of the conditioned variable itself. Some of the Z500 patterns obtained using RDA may be seen as perturbations of those obtained using PCA, but with major changes in the relative importance of patterns for surface field prediction.

The imprints of classical weather regimes on 10 m wind speed and T2m are illustrated in Figure 2 in Bloomfield et al. (2019). Although RDA surface field patterns in Figure 1 in this work and the surface responses of weather regimes in Figure 2 in Bloomfield et al. (2019) are not measured in the same units, they can still be compared assuming an equivalent multiplication factor to U100 and T2m imprints in Figure 1. Overall, it is conspicuous that the surface imprints of RDA patterns are stronger and more concentrated over Europe compared with the surface responses of classical weather regimes. The imprints of RDA patterns on U100 in Figure 1 show anomalous meridional and zonal dipoles which are originally absent in the surface responses of classical weather regimes on 10 m wind speed (Figure 2 in Bloomfield et al. (2019)). Although there are similarities in the first two imprints of weather regimes and RDA patterns on T2m, the center of the anomaly of the imprint corresponding to the first RDA pattern is shifted towards the southwest, while the imprint corresponding to the second RDA pattern shows a stronger dipole with a stretched northern anomaly center relative to the responses of weather regimes on T2m. The imprint of the third RDA pattern on T2m is significantly different from that of the Atlantic regime in Figure 2 of Bloomfield et al. (2019) and shows pronounced variations along with a tripole. The subsequent patterns are not shown in this work, but they present similar characteristics. Overall, the patterns in Figure 1 indicate stronger surface imprints of RDA patterns compared with weather regimes on both U100 and T2m.

Having understood the differences in surface imprints between weather regimes and RDA patterns, we now com-

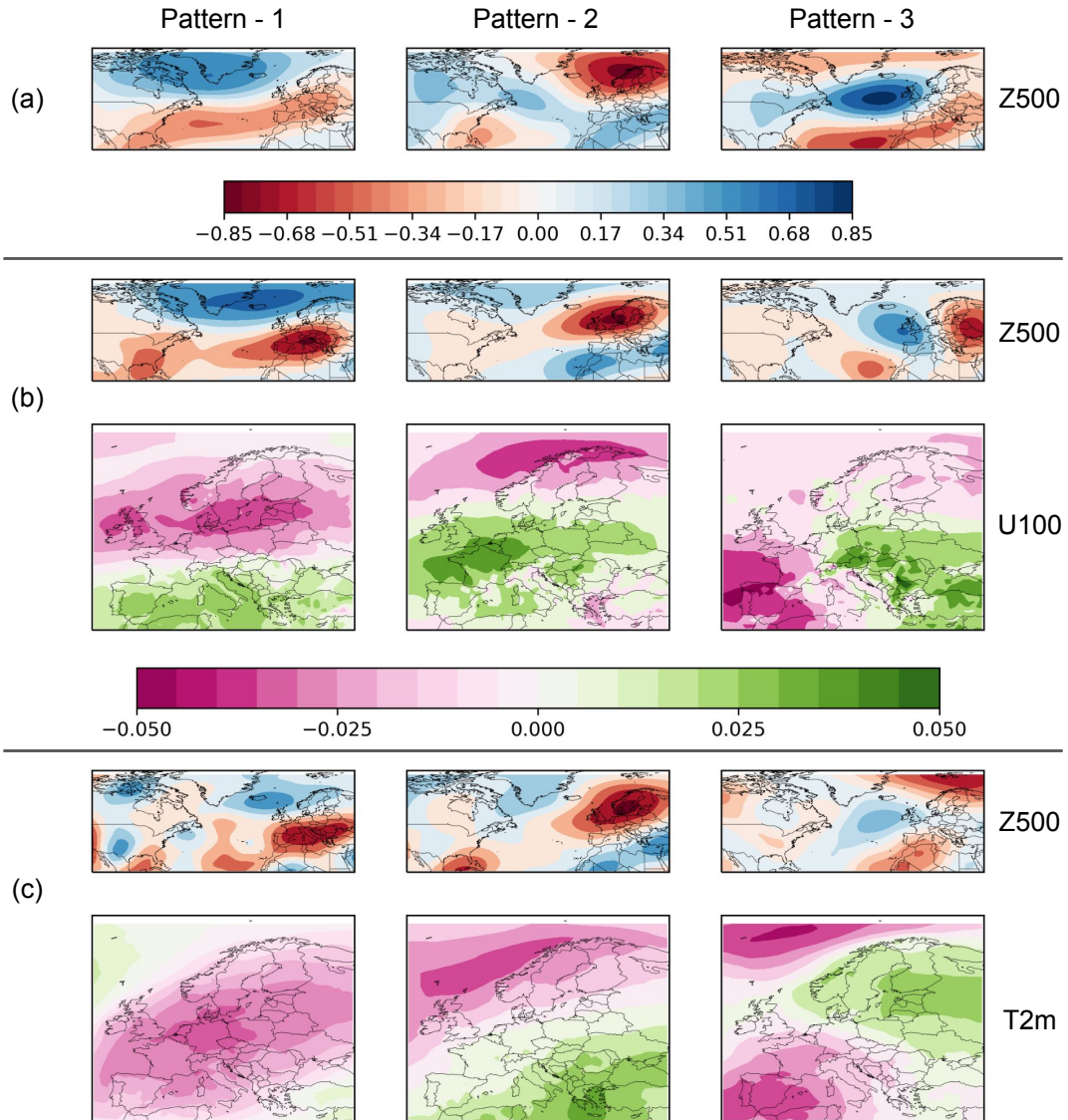


FIG. 1. (a) The first three patterns or Empirical Orthogonal Functions of Z500 anomalies over Euro-Atlantic computed through Principal Component Analysis. (b) The first three paired patterns of Redundancy Analysis of Z500 anomalies conditioned on U100 anomalies. (c) The first three paired patterns of Redundancy Analysis of Z500 anomalies conditioned on T2m anomalies. Top rows in (b) and (c) represent Z500 RDA patterns, and the corresponding bottom rows represent imprints of U100 and T2m, respectively. Please note that the color bars have no units, and that the signs are arbitrary. The patterns shown are unrotated.

pare the statistical explanatory power between PCA and RDA Z500 patterns, when used as predictors in a multilinear regression model, to accurately reconstruct surface fields. In Figure 2, we compare the performance of regression between Principal Component Regression (PCR) and RDA models. Particularly, we use coefficient of determination ( $R^2$ ) and root mean squared error (r.m.s.e.) of the linear fit between the predictor and the predictand as evaluation metrics. The  $R^2$  measures the proportion

of variation of the predictand that is accounted for by regression. Accordingly, the higher the  $R^2$ , the more is the predictand explained by the predictor. Kindly note that the  $R^2$  presented in Figure 2 shows the explained variance of individual grid points, while redundancy analysis maximizes the  $R^2$  averaged over the domain. In PCR, we first compute the PCs of the Euro-Atlantic Z500 anomaly field through Principal Component Analysis, and then use these PCs as predictors to predict U100 and T2m over Europe

via standard linear regression (Wilks 2019). In Figure 2, we retain the same number of PCs (i.e., all) for both PCR and RDA methods to facilitate comparison. From Figure 2, it is conspicuous that the  $R^2$  of RDA, with domain averages for U100 and T2m being 0.83 and 0.94, respectively, is substantially higher than that of PCR (domain averages of U100 and T2m being 0.46 and 0.56, respectively) for both the variables. Consequently, the r.m.s.e. of the linear fit between the predictor and the predictand of RDA, with domain averages for U100 and T2m being 0.53 and 0.46, respectively, is lower than that of PCR (domain averages of U100 and T2m being 0.96 and 1.27, respectively). The spatial variations of r.m.s.e. of U100 and T2m resemble that of the inter-annual variability of the respective variables (Appendix c). The  $R^2$  of RDA for T2m is relatively high compared with U100. Despite RDA models having higher  $R^2$ , the  $R^2$  for U100 drops to values below 0.5 over mountainous and other regions where the local effects are considerable. In general, using RDA Z500 patterns conditioned on the targeted predictand is advantageous over the use of Z500 patterns derived using PCA.

*b. How do the different types of ensemble predictions compare?*

In this section, we compare the skill of dynamical, statistical, hybrid, and climatological predictions in predicting U100 and T2m over Europe. To begin, we consider one case for illustrative purposes: the temporal evolution of ensemble Probability Density Functions (PDFs) of dynamical, statistical, hybrid, and climatological U100 predictions over southern Scandinavia initialized on 6 February 2017 is illustrated in Figure c), and the PDFs are computed for weekly means. The domain averaging is computed as the mean of the cosine of latitude weighted grid point values. Besides having a longer skill horizon of sub-seasonal U100 predictions compared to other European regions (Goutham et al. 2022), southern Scandinavia is one of the most important regions for the wind energy industry in Europe (WindEurope 2022). Hence, we consider southern Scandinavia for illustration purposes. This specific forecast (initiated on 6 February 2017) was chosen as it is qualitatively representative of the overall results. In Figure 3, the climatological predictions correspond to the same week and month of the year as the dynamical predictions but taken over each of the previous 35 years.

In week-3 in Figure 3, the PDF of the dynamical prediction ( $\mu = 7.03 \text{ m s}^{-1}$  and  $\sigma = 1.08 \text{ m s}^{-1}$ ) is closer to the observation ( $= 7.22 \text{ m s}^{-1}$ ), and therefore it appears to be more accurate than statistical prediction ( $\mu = 5.88 \text{ m s}^{-1}$  and  $\sigma = 0.76 \text{ m s}^{-1}$ ). However, dynamical predictions begin to converge towards their model climatology starting week-4. The statistical predictions are usually sharper compared with dynamical predictions at short lead times. This is attributed to slower evolution of large-scale fields

compared to surface fields (e.g., Buizza and Leutbecher 2015; Robertson and Vitart 2018). Beyond week-4, statistical predictions typically carry more valuable information relative to their dynamical counterparts and thus contribute greatly to hybrid prediction accuracy. To illustrate, the week-6 statistical prediction ( $\mu = 6.45 \text{ m s}^{-1}$  and  $\sigma = 0.82 \text{ m s}^{-1}$ ) is closer to observation ( $= 6.73 \text{ m s}^{-1}$ ) compared with dynamical prediction ( $\mu = 5.79 \text{ m s}^{-1}$  and  $\sigma = 0.99 \text{ m s}^{-1}$ ). The statistical predictions are not always perfect, and they are indeed only as good as the skill of large-scale fields in dynamical predictions. They fail when dynamical predictions fail, for instance when dynamical predictions are initialized during days closer to sudden stratospheric warming events (e.g., Gerber et al. 2009; Tripathi et al. 2015). For curious readers, some additional examples of comparison of different predictions are illustrated in the supplementary material. Overall, the PDFs in Figure 3 suggest that the hybrid predictions may be more accurate than either dynamical or statistical predictions beyond week-3.

We understood the behavior of different ensemble predictions for one particular forecast in Figure 3. In this section, we look at an overall assessment of U100 forecasts initiated in the boreal winter between December 2016 and February 2020 over southern Scandinavia. The comparison of the temporal evolution of Fair-CRPSS between dynamical, statistical, and hybrid U100 predictions over southern Scandinavia is shown in Figure 4. These violin plots are produced by aggregating the domain-averaged Fair-CRPSS of all the forecasts initiated in boreal winter between December 2016 and February 2020. In week-3, the ocean-atmosphere coupled dynamical predictions still carry important information about U100 over southern Scandinavia. The dynamical predictions, with a mean of FCRPSS of -0.0008 and a median of FCRPSS of 0.06, perform better than statistical predictions (mean = -0.10 and median = -0.04). The PDF of statistical predictions is heavily skewed towards negative values. The hybrid predictions, with a mean of 0.02 and a median of 0.07, perform better than either dynamical or statistical predictions. In week-3, dynamical predictions have a major contribution (compared with statistical predictions) to the improved skill of hybrid predictions. Contrary to dynamical predictions, the statistical predictions become more skillful with increasing lead time. Hence, statistical predictions contribute considerably to the improved skill of hybrid predictions at longer leads compared to shorter leads. The interquartile range (IQR) of hybrid predictions decreases with increasing lead time. This can be attributed to the decreasing IQR of statistical predictions with lead time. Beyond week-3, the improvements of hybrid predictions relative to their dynamical counterparts are statistically significant with  $p$ -values  $\leq 0.005$  based on Wilcoxon signed-rank test. Overall, the means and medians of FCRPSS of hybrid predictions, taking advantage of the strengths of the compo-

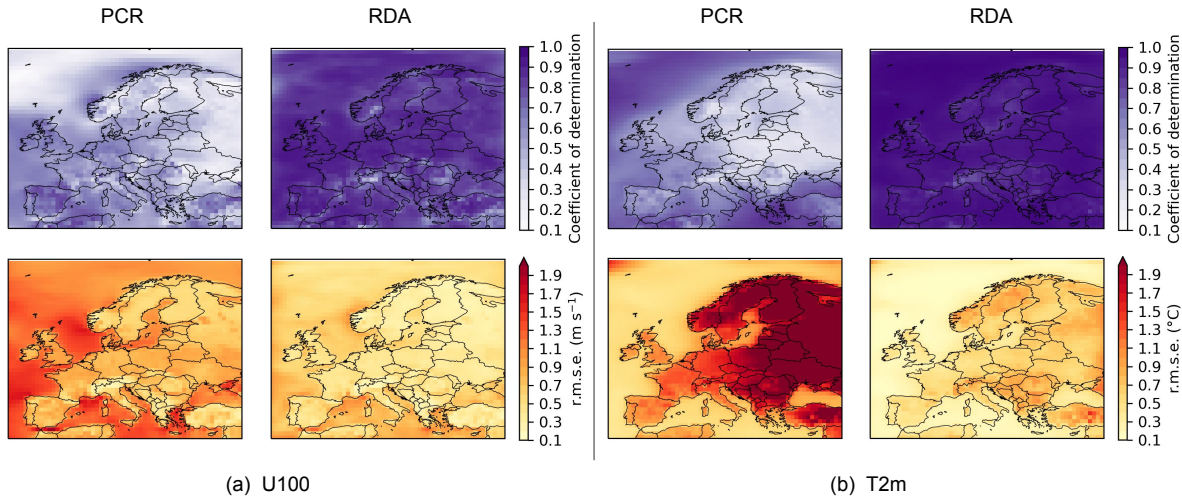


FIG. 2. Comparison of regression performance between Principal Component Regression (PCR) and Redundancy Analysis (RDA) models. Performance is measured using coefficient of determination ( $R^2$ ) of the linear fit between the predictor and the predictand, and root mean squared error (r.m.s.e.) between the predicted values and ERA5 reanalysis. The models are fit for weekly averages for the boreal winter between December 1999 and February 2016. (a) U100. (b) T2m.

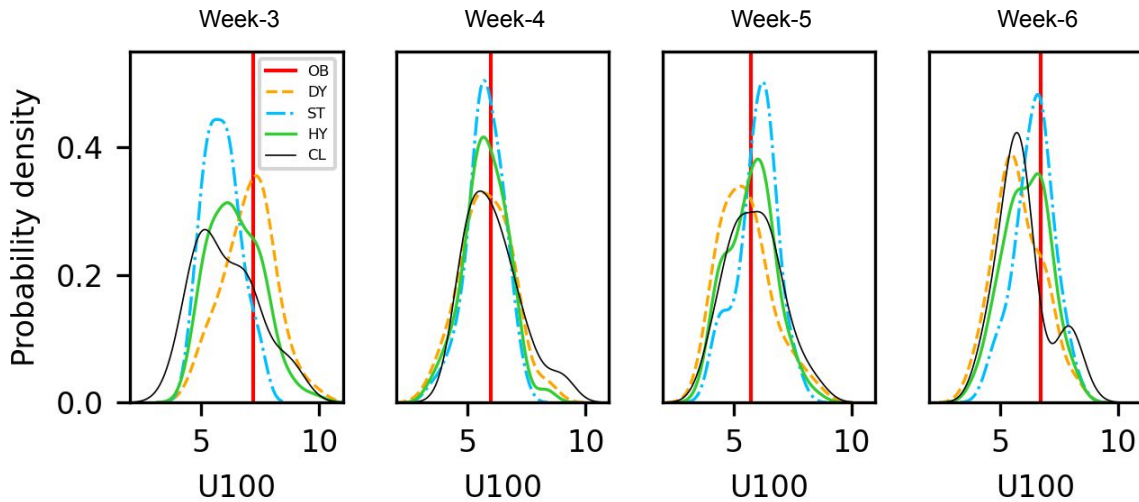


FIG. 3. Illustration of the temporal evolution of ensemble Probability Density Functions (PDFs) of dynamical (DY), statistical (ST), hybrid (HY), and climatological (CL) U100 predictions. The PDFs are computed as kernel-density estimates (Gaussian kernel) using ensemble members of weekly mean values averaged over southern Scandinavia ( $52.0^\circ - 61.0^\circ$  N,  $4.4^\circ - 19.0^\circ$  E). The grid point values are weighted by the cosine of their respective latitude before computing domain average. This illustration corresponds to dynamical predictions initialised on 6 February 2017. The red vertical line in each of the panels indicates the observed weekly mean (OB).

nent prediction systems, are both positive and higher than either dynamical or statistical predictions at all lead times.

For a more general assessment and to understand spatial variations of skill, we now compare the skill of dynamical and hybrid predictions at the scale of grid points over Europe. Figure 5 shows the comparison of the weekly evolution of mean-FCRPSS and FPSF between dynamical and hybrid U100 predictions across Europe. Although the hybrid prediction skill in week-3 is marginally poorer relative to that of dynamical predictions over southern Europe, the former has a relatively more positive mean-FCRPSS over northern Europe. Generally, the dynamical predictions,

lution of mean-FCRPSS and FPSF between dynamical and hybrid U100 predictions across Europe. Although the hybrid prediction skill in week-3 is marginally poorer relative to that of dynamical predictions over southern Europe, the former has a relatively more positive mean-FCRPSS over northern Europe. Generally, the dynamical predictions,



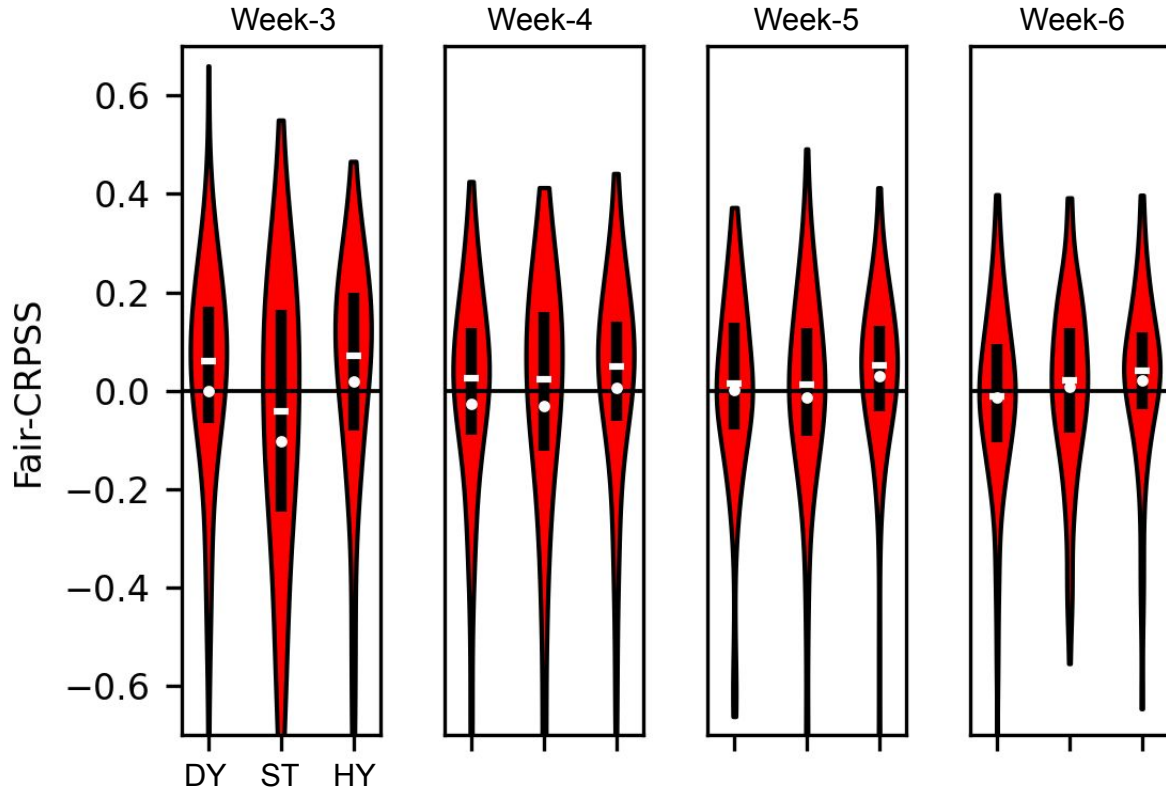


FIG. 4. Illustration of the temporal evolution of Fair-CRPSS of dynamical (DY), statistical (ST), and hybrid (HY) U100 predictions averaged over southern Scandinavia. In these standard violin plots, horizontal white dash indicate median, white circles indicate mean, black boxes indicate first and third quartiles, and black curves symmetric about the vertical (enclosing the red region) indicate the probability density of the Fair-CRPSS. The left-hand, central, and right-hand violin plots in each of the panels correspond to dynamical, statistical, and hybrid predictions, respectively. Values above zero indicate skillfulness of the respective predictions relative to climatology.

with an exception over and around the North Sea, are hardly skillful beyond week-3. The added value of the information from the slowly evolving large-scale fields through statistical predictions can be clearly noticed starting week-5. In week-6, the average of FPSF over Europe for dynamical and hybrid predictions are 49.6% and 54.1%, respectively. This indicates that the hybrid predictions outperform both the dynamical and statistical predictions (not shown) over a large part of Europe. Similar to the results presented in Ramon et al. (2021), the improvements brought in by the hybrid predictions are more pronounced over northern Europe than southern Europe. The number of patterns retained in statistical predictions that form a component of optimum U100 hybrid predictions increases slightly with lead time. The optimum U100 hybrid prediction skill is achieved when statistical predictions are produced using *eight to eleven* patterns on average, representing between 88% and 92% of the explained variance, depending on the lead time. As hybrid predictions are constructed by concatenating the ensemble members of dynamical and

statistical predictions, the poor hybrid prediction skill over southern Europe may be attributed to the poor skill of dynamical U100 predictions as well as low  $R^2$  of the linear fit between Z500 and U100 (Figure 2). Overall, hybrid predictions are more skillful than dynamical predictions at all lead times over a large part of Europe. Since a major proportion of the European wind farms are concentrated in and around the North Sea (WindEurope 2022), the wind energy industry could greatly benefit from improved hybrid predictions over this region.

Analogous to U100, Figure 6 compares the temporal evolution of mean-FCRPSS and FPSF between dynamical and hybrid T2m predictions over Europe. While hybrid predictions are typically more skillful than their dynamical counterparts over central, northern, and eastern Europe at all lead times, their skill is marginally degraded over southwestern Europe. Since we use a statistical model trained on the predictor and predictand *anomalies*, we notice the presence of cold biases in statistical T2m predictions attributed to the ongoing climate warming. This observation

12

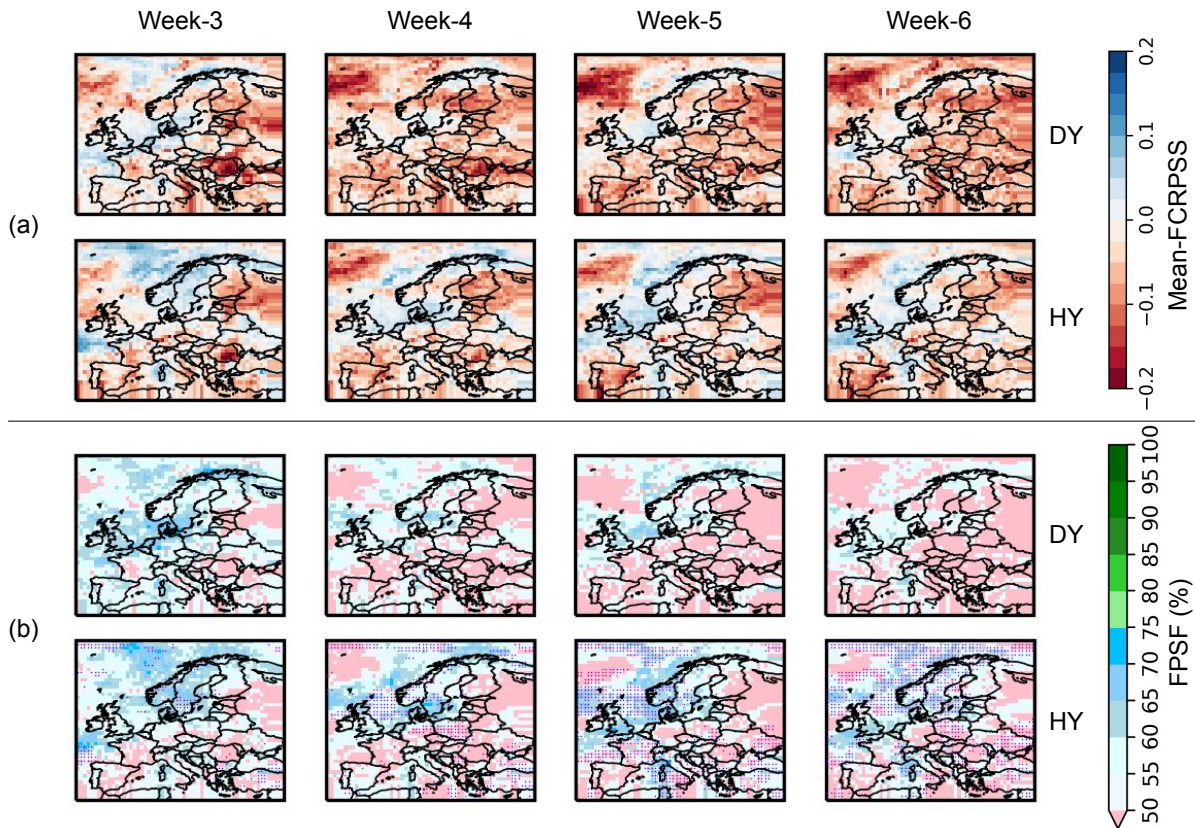


FIG. 5. Comparison of the temporal evolution of skill between dynamical (DY) and hybrid (HY) U100 predictions across Europe. (a) Mean-FCRPSS. (b) Fair-Proportion of Skillful Forecasts (FPSF). In (a) and (b), top rows correspond to dynamical predictions (DY) and bottom rows correspond to hybrid predictions (HY). Values above zero in (a) and above 50% in (b) indicate skillful predictions relative to climatology. Violet dots in hybrid predictions in (b) correspond to regions with statistically significant improvements at a significance level of  $p \leq 0.05$  based on Wilcoxon signed-rank test (Appendix b).

is consistent with the literature (e.g., Wilks 2013; Wilks and Livezey 2013; Wilks 2014). The presence of cold biases in statistical predictions translates to biased hybrid T2m predictions, and the poor hybrid prediction skill over certain regions in Figure 6 can be attributed to these biases. Additional post-processing of hybrid predictions may be required for them to be useful for practical applications, and this will be discussed in section 4c of the manuscript. Contrary to U100, the optimum T2m hybrid prediction skill is achieved when statistical predictions are produced using *three* patterns on average, representing about 75% of the explained variance. The number of patterns retained in statistical predictions that form a component of optimum T2m hybrid predictions is virtually insensitive to lead time. The differences in the number of retained patterns between U100 and T2m can be attributed to the complexity of the fields themselves. Overall, hybrid predictions benefit from both the skillful dynamical predictions of surface fields at shorter leads, and the longer skill horizon of large-scale fields and their statistical relationship with surface fields

at longer leads, and hence are usually more skillful than either dynamical or statistical predictions.

*c. Which forecast quality attribute(s) improves the hybrid prediction accuracy?*

The previous section has shown how the dynamical and statistical predictions complement each other to make the hybrid ensemble predictions more accurate than their components. In this section, we explore the differences in other forecast quality attributes such as association, reliability, resolution, and sharpness between dynamical and hybrid predictions, to understand the reasons for differences in accuracy between the two.

To understand the differences in association, we compare the temporal evolution of ACC between dynamical and hybrid predictions of U100 and T2m over Europe in Figure 7. The ACC of dynamical predictions of T2m is typically higher than that of U100 at all lead times. For U100 dynamical predictions, the ACC values drop below 0.4 starting week-4, whereas, for T2m, the ACC values drop below 0.4



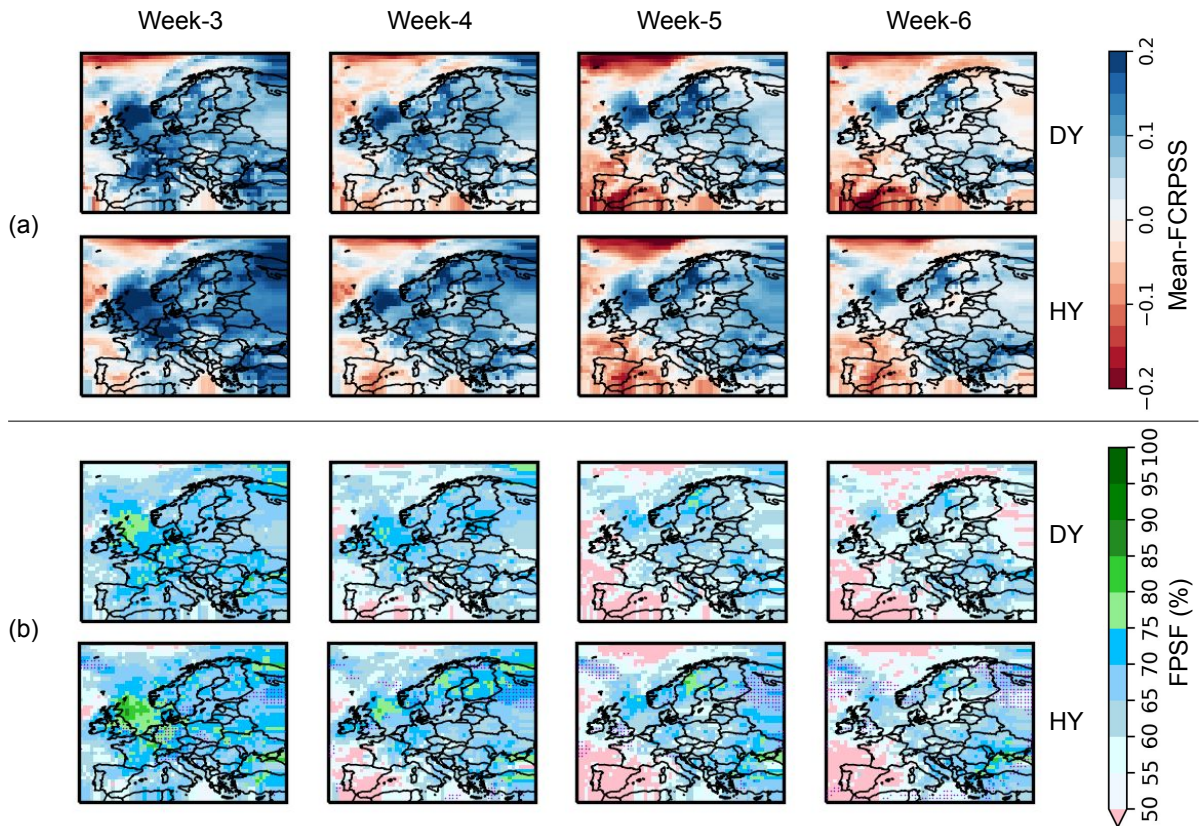


FIG. 6. As in Figure 5, but for T2m.

starting week-5. Overall, the differences in ACC between dynamical and hybrid predictions are marginal. The ACC of week-3 hybrid U100 predictions, relative to dynamical predictions, is lower over southern Europe and stronger over northern Europe. However, the differences in ACC between week-4 hybrid and dynamical U100 predictions are marginal over the European domain. There are only marginal differences in ACC between hybrid and dynamical T2m predictions over continental Europe in week-3 and week-4. Similar to Figures 5 and 6, the improvements in ACC brought in by hybrid predictions are noticeable starting week-5. Although the ACC remains poor (i.e.,  $\leq 0.4$ ) for U100 predictions starting week-5, the hybrid predictions marginally improve ACC over a large part of the domain. The ACC of T2m hybrid predictions is also marginally improved over more than two-thirds of continental Europe starting week-5.

We now compare the differences in reliability, resolution, and sharpness between dynamical and hybrid predictions with the help of reliability diagrams. We recall that reliability is a measure of calibration of the issued forecast probabilities. In a reliability diagram, the reliability component can be measured as the weighted average of

the squared difference between the points and the diagonal line. The number of forecasts in each bin is used as weights. The smaller the vertical distance between the points and the diagonal line, the more reliable are the predictions. In other words, perfectly reliable predictions have forecast probabilities essentially equal to observed frequencies, and hence all the points fall on the  $45^\circ$  diagonal line. The climatological line is the vertical or horizontal line drawn at the theoretical climatological probability of occurrence of the event considered (e.g., the climatological probability for a tercile is  $1/3$ ). Resolution measures the variations in the frequency of occurrence of an event as a function of the issued forecast probability. The resolution component can be measured as the weighted average of the squared difference between the points and the horizontal climatological line. The larger the vertical distance between the points and the horizontal climatological line, the higher the resolution. If the line connecting the points shows persistent offset from the  $45^\circ$  diagonal line, it indicates the presence of unconditional biases. Sharpness is a measure of the ability of forecasts to produce concentrated predictive distributions that are distinct from climatological probabilities. In a reliability diagram, the larger the hori-

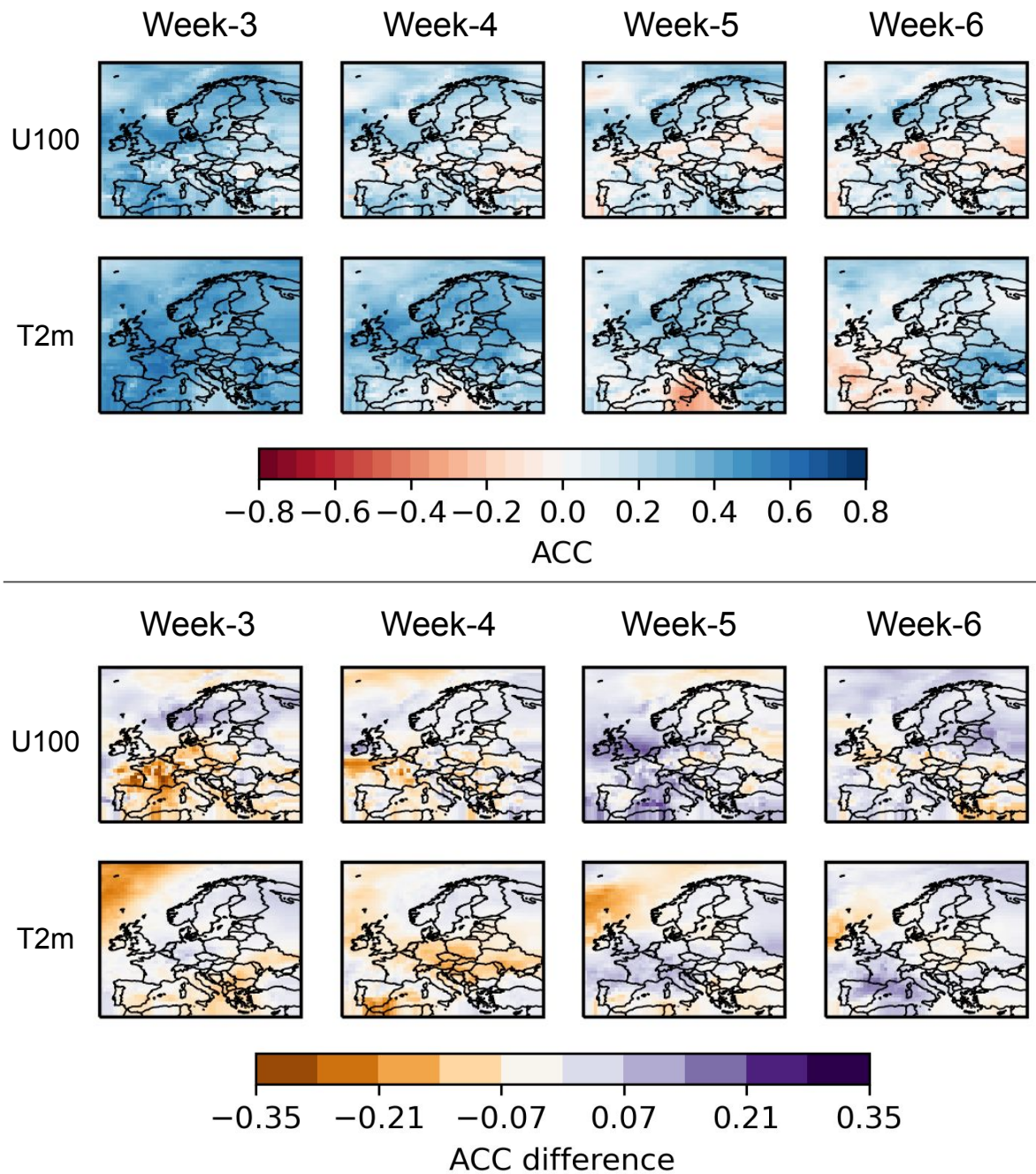


FIG. 7. (Top) Temporal evolution of Anomaly Correlation Coefficient (ACC) of dynamical predictions of U100 and T2m over Europe. (Bottom) Temporal evolution of the difference of ACC between hybrid and dynamical predictions (i.e.,  $ACC_{HY} - ACC_{DY}$ ) over Europe.

zontal distance between the climatological probability bin and the bin containing the maximum number of forecast instances, the sharper the predictions. The *no skill line* is the line located midway between the perfect reliability line and the horizontal climatological line. Accordingly,

the points located within the grey region bounded by the vertical climatological line and the *no skill line* contribute positively to skill. For a detailed description of reliability diagrams with examples, the reader is directed to Wilks (2019).

Figure 8 compares reliability diagrams between dynamical and hybrid models for upper and lower terciles of weekly mean U100 predictions for week-4 averaged across southern Scandinavia. We choose this particular domain and week for illustration purposes as the mean-FCRPSS of hybrid predictions is higher than that of dynamical predictions over this domain and during this week (Figure 5). For the upper tercile, hybrid predictions with a reliability component of 0.009 are more reliable than dynamical predictions (reliability = 0.016). Both the dynamical and hybrid predictions for upper tercile have a similar resolution ( $\sim 0.007$ ). However, for the lower tercile, both the reliability and resolution components of hybrid predictions are better than that of dynamical predictions. The difference in sharpness between dynamical and hybrid predictions is marginal. Reliability diagrams for the other lead times yield similar conclusions to the one obtained here: hybrid predictions are more reliable and have a better resolution than dynamical predictions (Figure A2 in Appendix d).

Figure 9 compares the reliability diagrams for upper and lower terciles of weekly mean T2m predictions for week-4 averaged across Germany ( $47.3^\circ - 55^\circ$  N,  $6.3^\circ - 15.4^\circ$  E) (Figure A1 in Appendix c) between dynamical and hybrid predictions. For the upper tercile, the reliability component of hybrid predictions (0.019) is higher than that of dynamical predictions (0.010), with both predictions having a similar resolution ( $\sim 0.02$ ). Although the reliability of upper tercile hybrid predictions looks similar to that of dynamical predictions at first sight, the hybrid predictions are in fact degraded due to the introduction of cold bias through the statistical model as a result of the warming climate. On the other hand for the lower tercile, the reliability component of hybrid predictions (0.013) is lower than that of dynamical predictions (0.018), with both predictions having a similar resolution ( $\sim 0.025$ ). The sharpness of the upper tercile hybrid predictions is marginally degraded relative to that of dynamical predictions.

We treat the cold bias of hybrid predictions by adjusting the warming trend through a simple procedure. Firstly, we compute the observed climatology for any given week and year under consideration by aggregating T2m of the same week and the two adjacent weeks over each of the previous 15 years from ERA5 reanalysis. Then, we assume the trend to be linear and fit a trend line to this climatological data. As we have chosen a 15-year period for climatology, the linearity assumption for the warming trend stays approximately valid (e.g., Wilks 2013; Wilks and Livezey 2013). Finally, we extrapolate the trend to the year under consideration and add it to statistical predictions. We then obtain trend-adjusted hybrid predictions ( $HY_{TrAd}$ ) by concatenating dynamical predictions to trend-adjusted statistical predictions. Adjusting for the trend in this way improves the reliability component of  $HY_{TrAd}$  for the upper (0.004) tercile without degrading the resolution. Since adjusting for the trend in this way shifts the entire distribution

to the right, it may degrade both reliability and resolution for the lower tercile. Nevertheless, trend-adjusted hybrid predictions show improved sharpness for both the terciles. The reliability diagrams for the remaining lead times is presented in Figure A3 (Appendix d). The intensity of the cold bias in the statistical model is different in different regions within Europe. There are other more sophisticated ways to deal with trend in a non-stationary climate, some of which are described in Wilks (2013), Wilks and Livezey (2013), and Wilks (2014). An alternative way to deal with the trend is to train on detrended T2m anomalies in the statistical model. Nevertheless, exploring the efficiencies of different trend adjustment methods will depend on the targeted application, and is hence beyond the scope of this research. Overall, the presented reliability diagrams show that the improved accuracy of hybrid predictions of surface fields relative to the corresponding dynamical predictions (Figures 4, 5 and 6) stems from improved reliability and resolution.

## 5. Discussion

### *a. How sensitive are the statistical predictions to the choice of predictor domain?*

The assessment of the previous section shows promising potential for extracting more skillful information from sub-seasonal predictions for surface variables using the methodology described in section 3. In this section, we discuss some of the choices and perspectives regarding this methodology.

We tested several predictor domains by varying size and geographical location to investigate the sensitivity of statistical prediction quality to the choice of domain. The predictor domain is tailored to describe the large-scale circulation, and more precisely features of this circulation that impact surface fields over Europe. Hence, it is logical to choose a predictor domain that is larger than the predictand domain. Given the mid-latitude circulation and dynamics (westerly flow, eastward traveling perturbations), it is expected that a domain shifted westward (i.e., upstream) should be best (Alonzo 2018). We recall that the predictor domain retained in this study is Euro-Atlantic ( $20^\circ - 80^\circ$  N,  $120^\circ$  W -  $40^\circ$  E). Displacing the predictor domain eastward (i.e., between  $90^\circ$  W -  $70^\circ$  E) without changing the size yields similar results to the retained Euro-Atlantic domain. However, choosing the predictor domain to be the same as that of the predictand domain (i.e.,  $34^\circ - 74^\circ$  N,  $13^\circ$  W -  $40^\circ$  E) considerably degrades statistical prediction quality. Shifting the predictand domain-sized predictor domain to the west between  $67^\circ$  W and  $14^\circ$  W marginally (but identifiably) degrades statistical prediction quality with respect to the case when the predictand domain-sized predictor domain is centered over the predictand domain. This confirms our hypothesis that the statistical model captures prominent



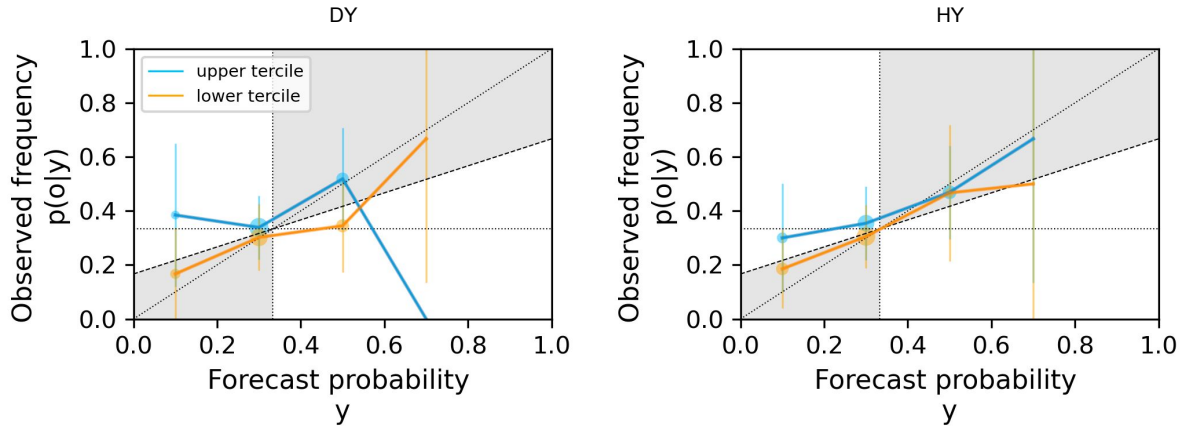


FIG. 8. Reliability diagrams for upper and lower terciles of weekly mean U100 predictions for week-4 averaged across southern Scandinavia (52.0° - 61.0° N, 4.4° - 19.0° E) (Figure A1 in Appendix c). The forecasts are stratified into five bins of equal width. The size of the points is proportional to the number of forecasts in the respective bins. The vertical bars refer to the 95% confidence intervals computed through the standard parametric approach by assuming a normal distribution for the underlying data (Machin et al. 2013). The vertical and horizontal dotted lines indicate the climatological tercile probabilities (theoretically, the value is 1/3) in the forecasts and observations, respectively. Perfectly reliable predictions fall on the dotted diagonal line (45°) connecting the points (0,0) and (1,1). The points located within the grey area contribute positively to skill. (DY) Dynamical predictions. (HY) Hybrid predictions.

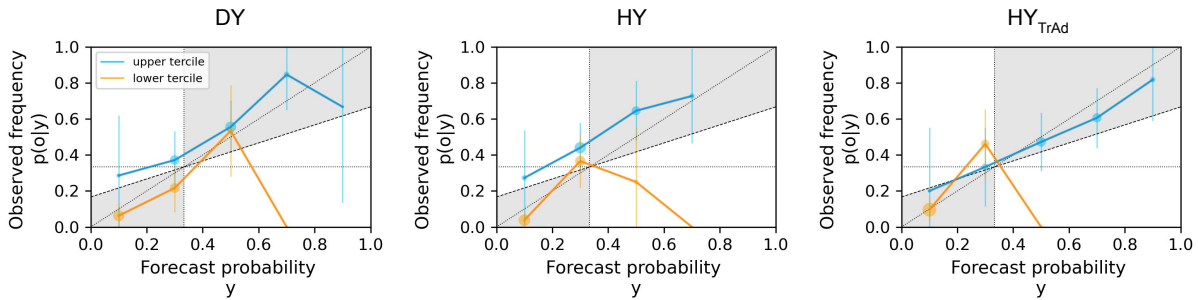


FIG. 9. Reliability diagrams for upper and lower terciles of weekly mean T2m predictions for week-4 averaged across Germany (47.3° - 55° N, 6.3° - 15.4° E) (Appendix c). The forecasts are stratified into five bins of equal width. The size of the points is proportional to the number of forecasts in the respective bins. The vertical bars refer to the 95% confidence intervals computed through the standard parametric approach by assuming a normal distribution for the underlying data (Machin et al. 2013). The vertical and horizontal dotted lines indicate the climatological tercile probabilities (theoretically, the value is 1/3) in the forecasts and observations, respectively. Perfectly reliable predictions fall on the dotted diagonal line (45°) connecting the points (0,0) and (1,1). The points located within the grey area contribute positively to skill. (DY) Dynamical predictions. (HY) Hybrid predictions. (HY<sub>TrAd</sub>) Trend-adjusted hybrid predictions.

information from the large-scale, mid-tropospheric westerly flow.

*b. How to further improve the hybrid prediction quality?*

In this study, we used a single predictor, i.e. Z500 over Euro-Atlantic, to improve the quality of surface field predictions. Other fields, tapping into other sources of sub-seasonal predictability (e.g., ocean, soil moisture, cryosphere), could be used in complement to Z500 to further improve the quality of surface field predictions (e.g., Seo et al. 2019; Domeisen et al. 2020). The redundancy analysis model can also be used as a tool to investigate physical as well as time-lag relationships between different

predictors and predictands, such as in the case of assessing impacts of MJO on extra-tropical weather (Zheng et al. 2018).

An additional way to improve hybrid predictions involves a more clever combination of ensembles from dynamical and statistical predictions. We recall that in this study, we concatenate dynamical and statistical predictions with each having an ensemble size of 50 to obtain a 100-member equally-weighted ensemble hybrid prediction. Nonetheless, combining ensembles through concatenation induces redundancy as some of the information that statistical prediction brings in may already be present in dynamical prediction. With the addition of other predictors,

the redundancy of ensemble members of hybrid predictions increases substantially.

As a first attempt to put proper weights on statistical and dynamical ensemble members of hybrid predictions, we classified statistical predictions into skillful or otherwise based on the values of observed redundancy PCs with respect to their climatological distribution when the dynamical predictions are initiated. In other words, we examined whether the observed redundancy PC values being in the lower or upper extreme quantiles with respect to their climatology at the time when dynamical predictions are initialized leads to the improved or degraded skill of statistical predictions. However, the results showed no detectable relationship between the two, and hence this path was not pursued further. A promising way forward here is through a Linear Inverse Model approach as realized in Albers and Newman (2021), and the authors plan to implement this in a future study.

There exist several alternative techniques to reasonably select ensemble members by minimizing redundancy. The most popular method for combining ensembles is Bayesian Model Averaging (BMA) (e.g., Schepen et al. 2012, 2014, 2016; Strazzo et al. 2019). BMA combines different models by giving different weights to the ensemble members from each model based on their performance. Another method for combining models that is gaining attention is the optimal transport distance or Wasserstein distance (e.g., Peyré and Cuturi 2019; Cumings-Menon and Shin 2020). The authors plan to investigate and compare these two methods in a future study.

## 6. Conclusions

With increasing decarbonization of the energy sector (IEA 2021), the energy industry requires accurate predictions of essential climate variables such as surface temperature and 100 m wind speed across a continuum of timescales. Having accurate predictions of essential climate variables on sub-seasonal timescales enables the energy industry to anticipate and prepare contingency plans in the face of anomalies in wind energy production and consumption (White et al. 2017). This calls for ways to improve the skill horizon of predictions of essential climate variables.

Surface variables such as wind speed and temperature are essential for many applications, yet in the forecast models, surface variables are not the most realistic. They are indeed strongly affected by small-scale, local features, and are heavily sensitive to parametrizations, which always introduce strong uncertainties. The skill horizon of predictions of surface variables is thus limited by errors in the representation of initial conditions, model formulations, and the use of restricted spatial resolution in sub-seasonal prediction models. Large-scale, low-frequency fields have the advantage of being more skillful than surface fields, and in

addition, they drive a large part of the variability of surface fields. In this study, we have proposed a novel methodology to improve predictions of surface variables by tapping into the large-scale, more reliable variables (e.g., Z500), and relating these to a surface variable of interest by training on the observationally derived historical data (i.e., ERA5 reanalysis). Generally, across Europe, weather regimes have been commonly used to provide a compact summary of the large-scale configuration of the atmosphere. For instance, Alonzo (2018) used weather regimes to summarize the large-scale atmospheric state and infer the likely surface wind speed distribution from these regimes using non-linear regression. Although weather regimes are powerful tools to anticipate surface conditions, the main limitations of the use of classical weather regimes for deducing surface fields are that these weather regimes represent large-scale atmospheric variability independently of the surface fields, and that each surface climate variable responds differently to the same weather regime. This calls for the development of new approaches to obtain large-scale spatial patterns of variability which take into account the variability of the targeted surface variable.

In this study, we have employed redundancy analysis to carry out a dimension reduction of the large-scale field (i.e., Z500). Redundancy analysis provides large-scale patterns specifically designed to capture the variability of a surface field of interest. We have compared the coefficients of determination between patterns obtained using Principal Component Analysis against those derived using Redundancy Analysis when used as predictors in a multi-linear regression model to reconstruct surface fields. We have then employed the relationship between patterns obtained using Redundancy Analysis and surface fields on the sub-seasonal dynamical predictions of patterns to obtain statistical probabilistic predictions of surface fields. Subsequently, we have combined statistical and dynamical predictions of surface fields through a simple concatenation of the respective ensemble members. From the results presented, the following conclusions can be drawn:

1. The large-scale patterns obtained using Redundancy Analysis better capture surface fields over Europe compared to patterns derived using Principal Component Analysis.
2. The added value of statistical predictions increases with lead time, and so does their contribution to the improved skill of hybrid predictions.
3. Combining dynamical and statistical predictions through a simple concatenation improves the skill of surface field predictions significantly over a large part of Europe at all lead times.
4. The improved accuracy of hybrid predictions relative to dynamical predictions stems from improved

reliability and resolution. No significant changes are observed in association and sharpness between dynamical and hybrid predictions.

5. The combination of dynamical and statistical predictions can certainly be improved. Depending on the initial state and/or the forecast evolution, one may have an a priori estimate of predictability, which could inform a more efficient combination of dynamical and statistical predictions.

The redundancy analysis model employed in this study can be used to identify spatial patterns of variability that impact surface conditions at a particular location of interest such as a wind farm. Wind farms, in addition to wind speed, require information about wind direction for operational purposes, e.g. to take into account the effect of wakes. In this regard, the redundancy analysis model can be employed for wind components separately. As a perspective, the redundancy analysis model could be deployed to identify spatial patterns of variability for other climate variables such as solar radiation and precipitation, and as well as for energy variables such as renewable energy production and consumption. The skill of statistical predictions realized in this study can be decomposed into two components: one, the skill of relevant large-scale Z500 patterns in the dynamical predictions, and two, the skill of regression. Since the patterns derived using Redundancy Analysis differ from those obtained using Principal Component Analysis, in terms of both spatial structure and explanatory power, it would be interesting to understand the differences in skill horizon between these patterns in dynamical predictions. As hybrid predictions developed in this study remain skillful even at a lead time of six weeks for both variables, it would be interesting to see their added value on seasonal timescales. These research questions along with the addition of other sources of predictability, and employment of more efficient ensemble selection are objectives for future studies.

*Acknowledgments.* Naveen Goutham would like to thank ANRT (Association Nationale de la Recherche et de la Technologie) for the CIFRE (Convention Industrielle de formation par la recherche) fellowship for his PhD.

This work contributes to the Energy4Climate Interdisciplinary Centre (E4C) of Institut Polytechnique de Paris and Ecole des Ponts ParisTech, supported by 3rd Programme d'Investissements d'Avenir [ANR-18-EUR-0006-02].

The authors acknowledge the data center ESPRI (Ensemble de services pour la recherche à l'Institut Pierre-Simon Laplace) for their help in storing and accessing the data. Finally, the authors would like to thank the editor and the two reviewers for carefully reviewing the work.

*Data availability statement.* The archived ECMWF extended-range forecasts and re-forecasts are published

under Creative Commons Attribution 4.0 International (CC BY 4.0). However, a data access fee may be applicable. For further information, kindly refer to <https://www.ecmwf.int/>. The ECMWF ERA5 reanalysis data is publicly available and can be accessed through the Climate Data Store of the Copernicus Climate Change Services upon registration.

The control of the ensemble should be treated as another indistinguishable ensemble member. However, due to the unavailability of the control member in the internal database of ESPRI (Ensemble de services pour la recherche à l'Institut Pierre-Simon Laplace) as a result of an unintentional man-made error, we had to use only the perturbed members.

## APPENDIX

### a. Forecast bias adjustment

The bias-adjusted ensemble member  $x_k^*$  for any forecast at a given lead time is

$$x_k^* = (x_k - \bar{x}_{cli}) \frac{\sigma_{ref}}{\sigma_{cli}} + \bar{o}_{ref}, \quad (A1)$$

where  $x_k$  is the raw member,  $\bar{x}_{cli}$  and  $\sigma_{cli}$  are the mean and standard deviation, respectively, of all the members of the re-forecast set corresponding to the forecast,  $\bar{o}_{ref}$  and  $\sigma_{ref}$  are the mean and standard deviation, respectively, of ERA5 reanalysis corresponding to the re-forecasts.

### b. Wilcoxon signed-rank test for statistical significance

Wilcoxon signed-rank test is a non-parametric hypothesis test that is used to investigate whether the considered data samples are derived from the same population or generating process (Wilcoxon 1945; Conover 1971; Wilks 2019). The test assesses for possible differences in location (i.e., rank) between members of a paired dataset. Here, the test statistic is based on ranks rather than numerical values of the data.

In this study, we consider a total of 103 forecasts initialized in the boreal winter between December 2016 and February 2020. We define the null hypothesis as "there is no difference in fair-CRPSS between dynamical and hybrid predictions", and the alternate hypothesis as "hybrid predictions have higher fair-CRPSS than dynamical predictions". We treat *ties* in the paired data using the method described in Pratt (1959).

### c. Inter-annual variability of U100 and T2m over Europe

The inter-annual variability of U100 and T2m over Europe is shown in Figure A1.

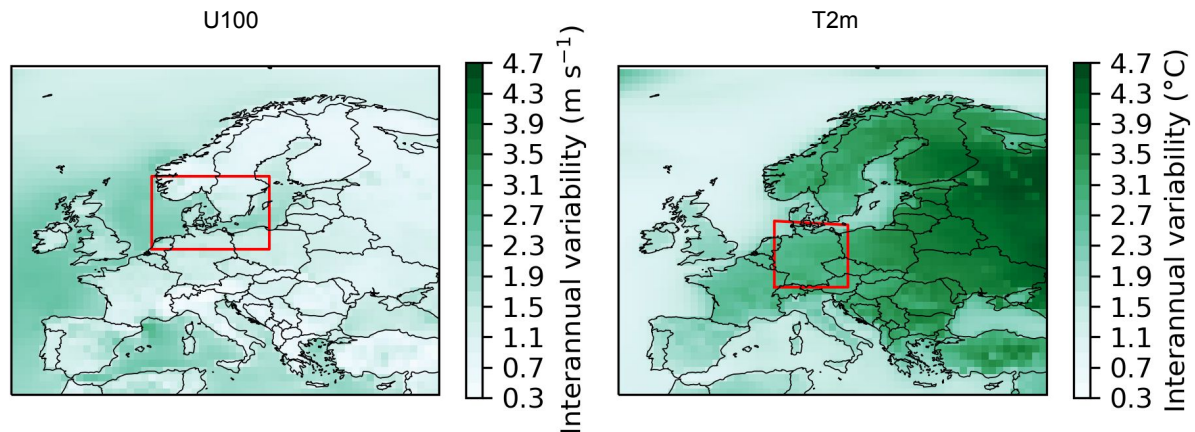


FIG. A1. Illustration of the inter-annual variability of boreal winter ERA5 reanalysis of 100 m wind speed (U100) and 2 m temperature (T2m) over Europe. The inter-annual variability is computed as the standard deviation of boreal winter weekly means between December 1999 and February 2016. The red colored rectangles in U100 and T2m correspond to southern Scandinavia (52.0° N - 61.0° N, 4.4° E - 19.0° E) and Germany (47.3° N - 55° N, 6.3°E - 15.4° E), respectively.

#### d. Additional reliability diagrams

The reliability diagrams for upper and lower terciles of weekly mean U100 and T2m predictions are shown in Figures A2 and A3, respectively.

#### References

- Albers, J. R., and M. Newman, 2021: Subseasonal predictability of the North Atlantic Oscillation. *Environmental Research Letters*, **16** (4), 044024, <https://doi.org/10.1088/1748-9326/abe781>.
- Alonzo, B., 2018: Seasonal forecasting of wind energy resource and production in France and associated risk. Ph.D. thesis, Université Paris-Saclay.
- Alonzo, B., H.-K. Ringkjøb, B. Jourdir, P. Drobinski, R. Plougonven, and P. Tankov, 2017: Modelling the variability of the wind energy resource on monthly and seasonal timescales. *Renewable energy*, **113**, 1434–1446.
- Baldwin, M. P., D. B. Stephenson, D. W. J. Thompson, T. J. Dunkerton, A. J. Charlton, and A. O'Neill, 2003: Stratospheric Memory and Skill of Extended-Range Weather Forecasts. *Science*, **301** (5633), 636–640, <https://doi.org/10.1126/science.1087143>, publisher: American Association for the Advancement of Science Section: Report.
- Benestad, R. E., D. Chen, and I. Hanssen-Bauer, 2008: *Empirical-statistical downscaling*. World Scientific Publishing Company.
- Bloomfield, H. C., D. J. Brayshaw, and A. J. Charlton-Perez, 2019: Characterizing the winter meteorological drivers of the European electricity system using targeted circulation types. *Meteor. Appl.*, **27** (1), e1858.
- Brune, S., J. Keller, and S. Wahl, 2021: Evaluation of wind speed estimates in reanalyses for wind energy applications. *Advances in Science and Research*, **18**, 115 – 126, <https://doi.org/10.5194/asr-18-115-2021>.
- Buizza, R., and M. Leutbecher, 2015: The forecast skill horizon. *Quart. J. Roy. Meteor. Soc.*, **141** (693), 3366–3382, <https://doi.org/10.1002/qj.2619>.
- Buizza, R., M. Leutbecher, and L. Isaksen, 2008: Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.: A journal of the atmospheric sciences*,

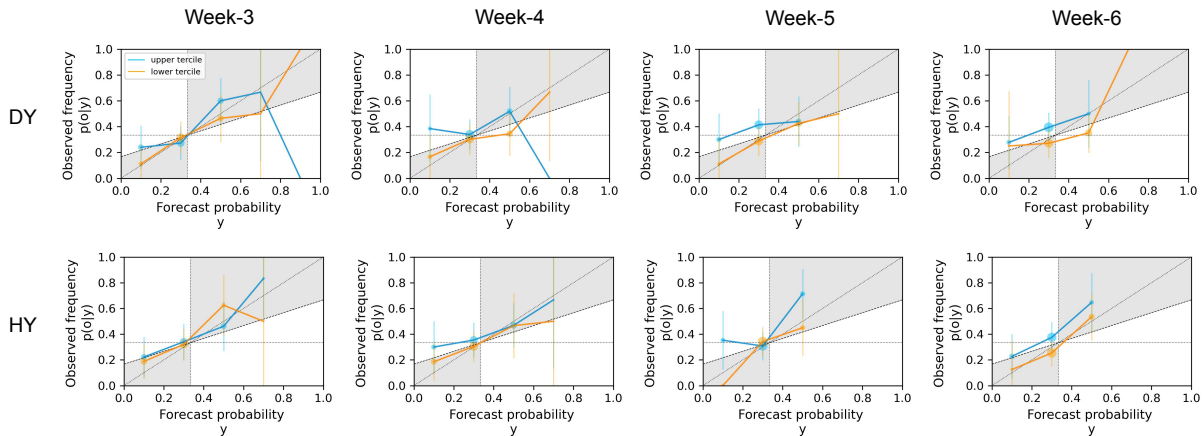


FIG. A2. As in Figure 8, but includes additional lead times.

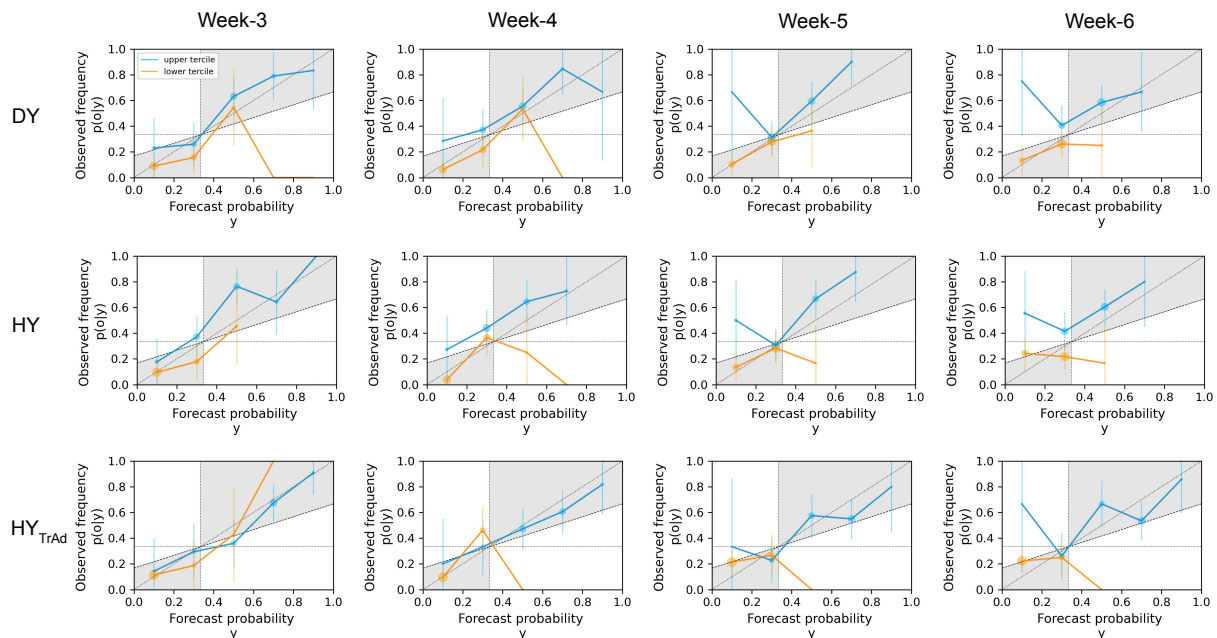


FIG. A3. As in Figure 9, but includes additional lead times.

*applied meteorology and physical oceanography*, **134** (637), 2051–2066.

Buizza, R., M. Leutbecher, and A. Thorpe, 2015: Living with the butterfly effect: A seamless view of predictability. *ECMWF newsletter*, **145**, 18–23.

Buizza, R., M. Milleer, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125** (560), 2887–2908, <https://doi.org/10.1002/qj.49712556006>.

Büeler, D., L. Ferranti, L. Magnusson, J. F. Quinting, and C. M. Grams, 2021: Year-round sub-seasonal forecast skill for Atlantic–European

weather regimes. *Quart. J. Roy. Meteor. Soc.*, **147** (741), 4283–4309, <https://doi.org/10.1002/qj.4178>.

Cheng, X., and J. M. Wallace, 1993: Cluster analysis of the northern hemisphere wintertime 500-hpa height field: Spatial patterns. *J. Atmos. Sci.*, **50** (16), 2674–2696.

Coelho, C. A. S., B. Brown, L. Wilson, M. Mittermaier, and B. Casati, 2019: Forecast verification for S2S timescales. *Sub-Seasonal to Seasonal Prediction: The Gap between Weather and Climate Forecasting*, A. Robertson, and F. Vitart, Eds., Elsevier, 337–361.

Conover, W., 1971: *Practical Nonparametric Statistics*. Wiley, URL <https://books.google.fr/books?id=Nv4YAAAAIAAJ>.



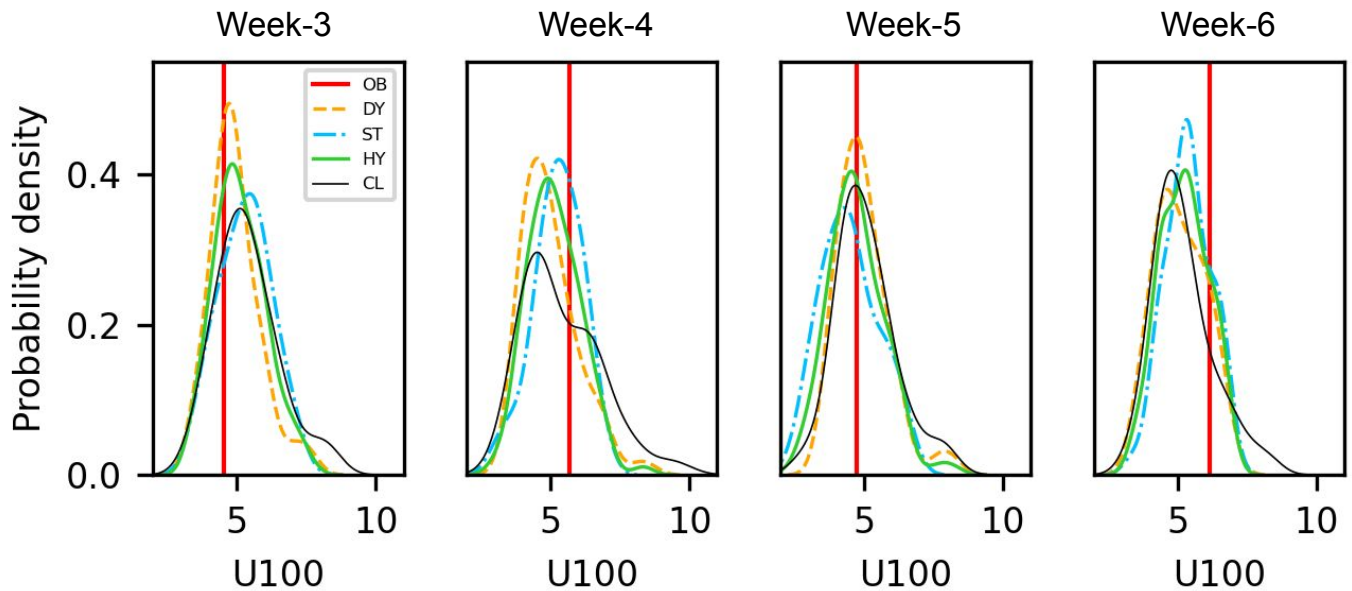
- Cummings-Menon, R., and M. Shin, 2020: Probability forecast combination via entropy regularized Wasserstein Distance. *Entropy*, **22** (9), URL <https://www.mdpi.com/1099-4300/22/9/929>.
- Diro, G. T., and H. Lin, 2020: Subseasonal Forecast Skill of Snow Water Equivalent and Its Link with Temperature in Selected SubX Models. *Wea. Forecasting*, **35** (1), 273–284, <https://doi.org/10.1175/WAF-D-19-0074.1>, publisher: Amer. Meteor. Soc. Section: Wea. Forecasting.
- Domeisen, D. I. V., and Coauthors, 2020: The role of the stratosphere in subseasonal to seasonal prediction: 2. predictability arising from stratosphere-troposphere coupling. *J. Geophys. Res.: Atmospheres*, **125** (2), e2019JD030923, <https://doi.org/10.1029/2019JD030923>.
- Dörenkämper, M., and Coauthors, 2020: The making of the new European wind atlas – part 2: Production and evaluation. *Geoscientific Model Development*, **13** (10), 5079–5102, <https://doi.org/10.5194/gmd-13-5079-2020>.
- Dorrington, J., I. Finney, T. Palmer, and A. Weisheimer, 2020: Beyond skill scores: exploring sub-seasonal forecast value through a case-study of French month-ahead energy prediction. *Quart. J. Roy. Meteor. Soc.*, **146** (733), 3623–3637, <https://doi.org/10.1002/qj.3863>.
- Ferro, C. a. T., 2014: Fair scores for ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, **140** (683), 1917–1923, <https://doi.org/10.1002/qj.2270>.
- Fu, X., B. Wang, D. E. Waliser, and L. Tao, 2007: Impact of Atmosphere–Ocean Coupling on the Predictability of Monsoon Intraseasonal Oscillations. *J. Atmos. Sci.*, **64** (1), 157–174, <https://doi.org/10.1175/JAS3830.1>, publisher: Amer. Meteor. Soc. Section: J. Atmos. Sci.
- Garrido-Perez, J. M., C. Ordóñez, D. Barriopedro, R. García-Herrera, and D. Paredes, 2020: Impact of weather regimes on wind power variability in western Europe. *Applied Energy*, **264**, 114 731, <https://doi.org/10.1016/j.apenergy.2020.114731>.
- Gerber, E. P., C. Orbe, and L. M. Polvani, 2009: Stratospheric influence on the tropospheric circulation revealed by idealized ensemble forecasts. *Geophys. Res. Lett.*, **36** (24), <https://doi.org/10.1029/2009GL040913>.
- Gneiting, T., and A. E. Raftery, 2007: Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, **102** (477), 359–378, <https://doi.org/10.1198/016214506000001437>.
- Goutham, N., R. Plougonven, H. Omrani, S. Parey, P. Tankov, A. Tantet, P. Hitchcock, and P. Drobinski, 2022: How skillful are the European sub-seasonal predictions of wind speed and surface temperature? *Mon. Wea. Rev.*, <https://doi.org/10.1175/MWR-D-21-0207.1>.
- Goutham, N., and Coauthors, 2021: Using machine-learning methods to improve surface wind speed from the outputs of a Numerical Weather Prediction model. *Boundary-Layer Meteorology*, **179** (1), 133–161, <https://doi.org/10.1007/s10546-020-00586-x>.
- Grams, C. M., R. Beerli, S. Pfenniger, I. Staffell, and H. Wernli, 2017: Balancing Europe’s wind-power output through spatial deployment informed by weather regimes. *Nature Climate Change*, **7** (8), 557–562, <https://doi.org/10.1038/nclimate3338>.
- Hersbach, H., 2000: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Wea. Forecasting*, **15** (5), 559 – 570, [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2), place: Boston MA, USA Publisher: Amer. Meteor. Soc.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146** (730), 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hewitson, B., and R. G. Crane, 1996: Climate downscaling: techniques and application. *Climate Research*, **7**, 85–95.
- Hoskins, B., 2012: Predictability Beyond the Deterministic Limit. World Meteorological Organization, URL <https://public.wmo.int/en/bulletin/predictability-beyond-deterministic-limit>.
- IEA, 2020: European Union 2020. Tech. rep., Paris, France. URL <https://www.iea.org/reports/european-union-2020>.
- IEA, 2021: Net Zero by 2050. Tech. rep., Paris, France. URL <https://www.iea.org/reports/net-zero-by-2050>.
- Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud, 2010: Ensemble of data assimilations at ECMWF. (636), 45, URL <https://www.ecmwf.int/node/10125>, publisher: ECMWF.
- Jifan, C., 1989: Predictability of the atmosphere. *Advances in Atmospheric Sciences*, **6** (3), 335–346, <https://doi.org/10.1007/BF02661539>.
- Johannsen, F., S. Ermida, J. P. A. Martins, I. F. Trigo, M. Nogueira, and E. Dutra, 2019: Cold bias of ERA5 summertime daily maximum land surface temperature over Iberian peninsula. *Remote Sensing*, **11** (21), URL <https://www.mdpi.com/2072-4292/11/21/2570>.
- Jolliffe, I., and D. Stephenson, 2003: *Forecast Verification: A Practitioner’s Guide in Atmospheric Science*. Wiley, URL <https://books.google.cm/books?id=cjS9kK8IWBwC>.
- Jonek-Kowalska, I., 2022: Towards the reduction of co2 emissions. paths of pro-ecological transformation of energy mixes in European countries with an above-average share of coal in energy consumption. *Resources Policy*, **77**, 102 701, <https://doi.org/10.1016/j.resourpol.2022.102701>.
- Jones, C., D. E. Waliser, K. M. Lau, and W. Stern, 2004a: Global Occurrences of Extreme Precipitation and the Madden–Julian Oscillation: Observations and Predictability. *J. Climate*, **17** (23), 4575–4589, <https://doi.org/10.1175/3238.1>, publisher: Amer. Meteor. Soc. Section: J. Climate.
- Jones, C., D. E. Waliser, K. M. Lau, and W. Stern, 2004b: The Madden–Julian Oscillation and Its Impact on Northern Hemisphere Weather Predictability. *Mon. Wea. Rev.*, **132** (6), 1462–1471, [https://doi.org/10.1175/1520-0493\(2004\)132<1462:TMOAII>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1462:TMOAII>2.0.CO;2), publisher: Amer. Meteor. Soc. Section: Mon. Wea. Rev.
- Jourdier, B., 2020: Evaluation of ERA5, MERRA-2, COSMO-REA6, NEWA and AROME to simulate wind power production over France. *Advances in Science and Research*, Copernicus GmbH, Vol. 17, 63–77, <https://doi.org/10.5194/asr-17-63-2020>.
- Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press.
- Koster, R. D., and Coauthors, 2011: The Second Phase of the Global Land–Atmosphere Coupling Experiment: Soil Moisture Contributions to Subseasonal Forecast Skill. *J. Hydrometeor.*, **12** (5), 805–822, <https://doi.org/10.1175/2011JHM1365.1>.
- Leutbecher, M., 2005: On ensemble prediction using singular vectors started from forecasts. (462), 11, URL <https://www.ecmwf.int/node/10732>.

- Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *Journal of computational physics*, **227** (7), 3515–3539.
- Leutbecher, M., and Coauthors, 2016: Model uncertainty representations in the IFS. *ECMWF/WWRP Workshop: Model Uncertainty*, ECMWF, Reading, URL <https://www.ecmwf.int/node/16369>.
- Lin, H., and Z. Wu, 2011: Contribution of the Autumn Tibetan Plateau Snow Cover to Seasonal Prediction of North American Winter Temperature. *J. Climate*, **24** (11), 2801–2813, <https://doi.org/10.1175/2010JCLI3889.1>, publisher: Amer. Meteor. Soc. Section: J. Climate.
- Liobikienė, G., and M. Butkus, 2017: The European union possibilities to achieve targets of Europe 2020 and Paris agreement climate policy. *Renewable Energy*, **106**, 298–309, <https://doi.org/10.1016/j.renene.2017.01.036>.
- Lledó, L., and F. J. Doblas-Reyes, 2020: Predicting Daily Mean Wind Speed in Europe Weeks ahead from MJO Status. *Mon. Wea. Rev.*, **148** (8), 3413–3426, <https://doi.org/10.1175/MWR-D-19-0328.1>.
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20** (2), 130–141.
- Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17** (3), 321–333, <https://doi.org/10.1111/j.2153-3490.1965.tb01424.x>.
- Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34** (6), 505–513.
- Lynch, K. J., D. J. Brayshaw, and A. Charlton-Perez, 2014: Verification of European Subseasonal Wind Speed Forecasts. *Mon. Wea. Rev.*, **142**, 13.
- Machin, D., T. Bryant, D. Altman, and M. Gardner, 2013: *Statistics with Confidence: Confidence Intervals and Statistical Guidelines*. Wiley, URL <https://books.google.fr/books?id=HmnIBAAQBAJ>.
- Manzanas, R., J. Gutiérrez, J. Fernández, E. van Meijgaard, S. Calmanti, M. Magariño, A. Cofiño, and S. Herrera, 2018: Dynamical and statistical downscaling of seasonal temperature forecasts in Europe: Added value for user applications. *Climate Services*, **9**, 44–56, <https://doi.org/10.1016/j.cliser.2017.06.004>, climate services in practice: what we learnt from EUPORIAS.
- Manzanas, R., J. M. Gutiérrez, J. Bhend, S. Hemri, F. J. Doblas-Reyes, V. Torralba, E. Penabaz, and A. Brookshaw, 2019: Bias adjustment and ensemble recalibration methods for seasonal forecasting: a comprehensive intercomparison using the C3S dataset. *Climate Dyn.*, **53** (3–4), 1287–1305, <https://doi.org/10.1007/s00382-019-04640-4>.
- Matheson, J. E., and R. L. Winkler, 1976: Scoring Rules for Continuous Probability Distributions. *Management Science*, **22** (10), 1087–1096, URL [https://econpapers.repec.org/article/inmormnsc/v\\_3a22\\_3ay\\_3a1976\\_3ai\\_3a10\\_3ap\\_3a1087-1096.htm](https://econpapers.repec.org/article/inmormnsc/v_3a22_3ay_3a1976_3ai_3a10_3ap_3a1087-1096.htm).
- Molina, M. O., C. Gutiérrez, and E. Sánchez, 2021: Comparison of ERA5 surface wind speed climatologies over Europe with observations from the HadISD dataset. *Int. J. Climatol.*, **41** (10), 4864–4878, <https://doi.org/10.1002/joc.7103>.
- Monhart, S., C. Spirig, J. Bhend, K. Bogner, C. Schär, and M. A. Liniger, 2018: Skill of Subseasonal Forecasts in Europe: Effect of Bias Correction and Downscaling Using Surface Observations. *J. Geophys. Res.: Atmospheres*, <https://doi.org/10.1029/2017JD027923>.
- Murcia, J. P., M. J. Koivisto, G. Luzia, B. T. Olsen, A. N. Hahmann, P. E. Sørensen, and M. Als, 2022: Validation of European-scale simulated wind speed and wind generation time series. *Applied Energy*, **305**, 117794, <https://doi.org/10.1016/j.apenergy.2021.117794>.
- Namias, J., 1952: The Annual Course of Month-to-Month Persistence in Climatic Anomalies. *Bull. Amer. Meteor. Soc.*, **33** (7), 279–285, <https://doi.org/10.1175/1520-0477-33.7.279>.
- Orsolini, Y. J., R. Senan, G. Balsamo, F. J. Doblas-Reyes, F. Vitart, A. Weisheimer, A. Carrasco, and R. E. Benestad, 2013: Impact of snow initialization on sub-seasonal forecasts. *Climate Dyn.*, **41** (7), 1969–1982, <https://doi.org/10.1007/s00382-013-1782-0>.
- Palmer, T. N., 2012: Towards the probabilistic Earth-system simulator: a vision for the future of climate and weather prediction. *Quart. J. Roy. Meteor. Soc.*, **138** (665), 841–861, <https://doi.org/10.1002/qj.1923>.
- Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. (598), 42, URL <https://www.ecmwf.int/node/11577>, publisher: ECMWF.
- Peyré, G., and M. Cuturi, 2019: *Computational Optimal Transport: With Applications to Data Science*. Foundations and trends in machine learning, Now Publishers, URL <https://books.google.fr/books?id=J0BiwgEACAAJ>.
- Plaut, G., and E. Simonnet, 2001: Large-scale circulation classification, weather regimes, and local climate over France, the Alps and Western Europe. *Climate Research*, **17** (3), 303–324.
- Pratt, J. W., 1959: Remarks on zeros and ties in the Wilcoxon Signed Rank procedures. *Journal of the American Statistical Association*, **54** (287), 655–667, <https://doi.org/10.1080/01621459.1959.10501526>.
- Prodhomme, C., F. Doblas-Reyes, O. Bellprat, and E. Dutra, 2016: Impact of land-surface initialization on sub-seasonal to seasonal forecasts over Europe. *Climate Dyn.*, **47** (3), 919–935, <https://doi.org/10.1007/s00382-015-2879-4>.
- Ramon, J., L. Lledó, P.-A. Bretonnière, M. Samsó, and F. J. Doblas-Reyes, 2021: A perfect prognosis downscaling methodology for seasonal prediction of local-scale wind speeds. *Environmental Research Letters*, **16** (5), 054010, <https://doi.org/10.1088/1748-9326/abe491>.
- Ramon, J., L. Lledó, V. Torralba, A. Soret, and F. J. Doblas-Reyes, 2019: What global reanalysis best represents near-surface winds? *Quart. J. Roy. Meteor. Soc.*, **145** (724), 3236–3251, <https://doi.org/10.1002/qj.3616>.
- Raoult, B., C. Bergeron, A. L. Alós, J.-N. Thépaut, and D. Dee, 2017: Climate service develops user-friendly data store. ECMWF, Shinfield Park, Reading, URL <https://www.ecmwf.int/en/newsletter/151/meteorology/climate-service-develops-user-friendly-data-store>, newsletter 151.
- Robertson, A., and F. Vitart, 2018: *Sub-seasonal to Seasonal Prediction: The Gap Between Weather and Climate Forecasting*. Elsevier.
- Sanders, F., 1963: On Subjective Probability Forecasting. *J. Appl. Meteor. Climatol.*, **2** (2), 191–201, [https://doi.org/10.1175/1520-0450\(1963\)002<0191:OSPF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1963)002<0191:OSPF>2.0.CO;2).
- Schepen, A., Q. J. Wang, and Y. Everingham, 2016: Calibration, bridging, and merging to improve GCM seasonal temperature forecasts in Australia. *Mon. Wea. Rev.*, **144** (6), 2421 – 2441, <https://doi.org/10.1175/MWR-D-15-0384.1>.

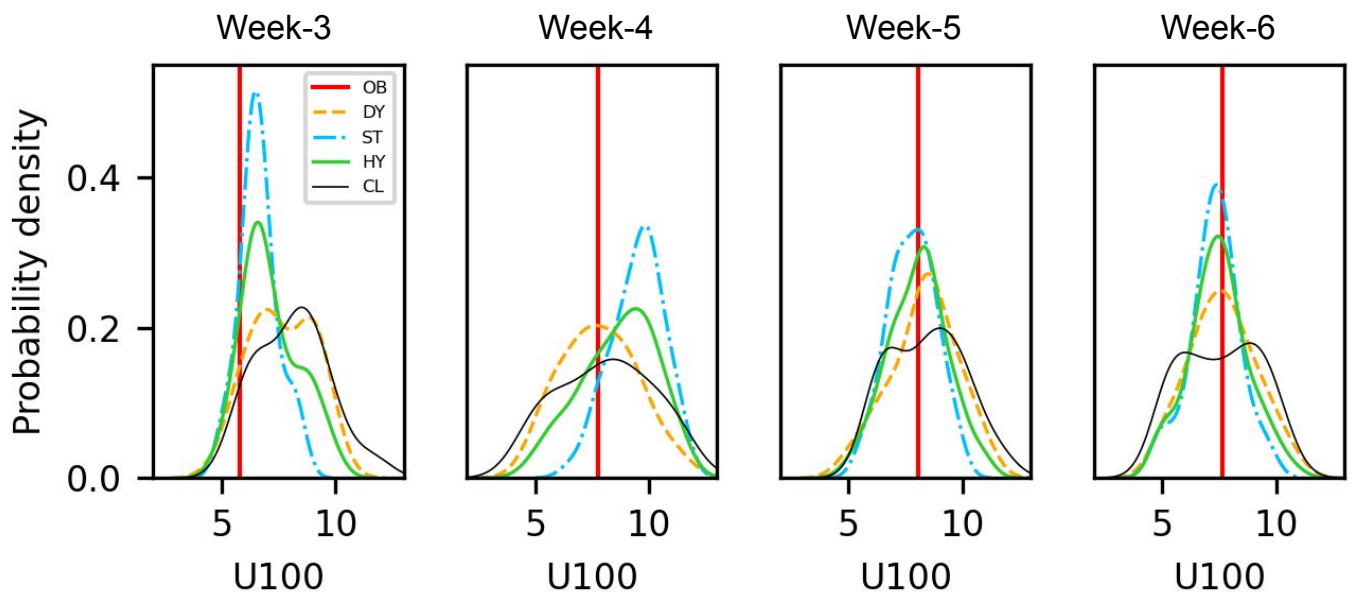
- Schepen, A., Q. J. Wang, and D. E. Robertson, 2012: Combining the strengths of statistical and dynamical modeling approaches for forecasting Australian seasonal rainfall. *J. Geophys. Res.: Atmospheres*, **117** (D20), <https://doi.org/10.1029/2012JD018011>.
- Schepen, A., Q. J. Wang, and D. E. Robertson, 2014: Seasonal forecasts of Australian rainfall through calibration and bridging of coupled GCM outputs. *Mon. Wea. Rev.*, **142** (5), 1758 – 1770, <https://doi.org/10.1175/MWR-D-13-00248.1>.
- Schwartz, C., and C. I. Garfinkel, 2020: Troposphere-stratosphere coupling in subseasonal-to-seasonal models and its importance for a realistic extratropical response to the Madden-Julian Oscillation. *J. Geophys. Res.: Atmospheres*, **125** (10), e2019JD032043, <https://doi.org/10.1029/2019JD032043>.
- Seo, E., and Coauthors, 2019: Impact of soil moisture initialization on boreal summer subseasonal forecasts: mid-latitude surface air temperature and heat wave events. *Climate Dyn.*, **52** (3), 1695–1709, <https://doi.org/10.1007/s00382-018-4221-4>.
- Simmons, A., and Coauthors, 2021: Low frequency variability and trends in surface air temperature and humidity from ERA5 and other datasets. Tech. Rep. 881, ECMWF. <https://doi.org/10.21957/ly5vbtbfd>.
- Sobolowski, S., G. Gong, and M. Ting, 2010: Modeled Climate State and Dynamic Responses to Anomalous North American Snow Cover. *J. Climate*, **23** (3), 785–799, <https://doi.org/10.1175/2009JCLI3219.1>, publisher: Amer. Meteor. Soc. Section: J. Climate.
- Strazzo, S., D. C. Collins, A. Schepen, Q. J. Wang, E. Becker, and L. Jia, 2019: Application of a hybrid statistical–dynamical system to seasonal prediction of North American temperature and precipitation. *Mon. Wea. Rev.*, **147** (2), 607 – 625, <https://doi.org/10.1175/MWR-D-18-0156.1>.
- Subramanian, A. C., and Coauthors, 2019: Ocean observations to improve our understanding, modeling, and forecasting of subseasonal-to-seasonal variability. *Frontiers in Marine Science*, **6**, <https://doi.org/10.3389/fmars.2019.00427>.
- Tarek, M., F. P. Brissette, and R. Arsenault, 2019: Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modeling over North-America. preprint, Catchment hydrology/Modelling approaches. <https://doi.org/10.5194/hess-2019-316>.
- Tippett, M. K., T. DelSole, S. J. Mason, and A. G. Barnston, 2008: Regression-Based Methods for Finding Coupled Patterns. *J. Climate*, **21** (17), 4384–4398, <https://doi.org/10.1175/2008JCLI2150.1>.
- Torralba, V., F. J. Doblas-Reyes, D. MacLeod, I. Christel, and M. Davis, 2017: Seasonal Climate Prediction: A New Source of Information for the Management of Wind Energy Resources. *J. Appl. Meteor. Climatol.*, **56** (5), 1231–1247, <https://doi.org/10.1175/JAMC-D-16-0204.1>.
- Toth, Z., and R. Buizza, 2019: Chapter 2 - weather forecasting: What sets the forecast skill horizon? *Sub-Seasonal to Seasonal Prediction*, A. W. Robertson, and F. Vitart, Eds., Elsevier, 17–45, <https://doi.org/10.1016/B978-0-12-811714-9.00002-4>.
- Tripathi, O. P., and Coauthors, 2015: The predictability of the extratropical stratosphere on monthly time-scales and its impact on the skill of tropospheric forecasts. *Quart. J. Roy. Meteor. Soc.*, **141** (689), 987–1003, <https://doi.org/10.1002/qj.2432>.
- Unger, D. A., 1985: A method to estimate the Continuous Ranked Probability Score. *Conference on Probability and Statistics in Atmospheric Sciences*, **9th**, 206–213, URL [https://jglobal.jst.go.jp/en/detail?JGLOBAL\\_ID=200902092398162270](https://jglobal.jst.go.jp/en/detail?JGLOBAL_ID=200902092398162270).
- van den Hurk, B., F. Doblas-Reyes, G. Balsamo, R. D. Koster, S. I. Seneviratne, and H. Camargo, 2012: Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe. *Climate Dyn.*, **38** (1), 349–362, <https://doi.org/10.1007/s00382-010-0956-2>.
- van der Wiel, K., H. C. Bloomfield, R. W. Lee, L. P. Stoop, R. Blackport, J. A. Screen, and F. M. Selten, 2019: The influence of weather regimes on European renewable energy production and demand. *Environmental Research Letters*, **14** (9), 094010.
- Velikou, K., G. Lazoglou, K. Tolika, and C. Anagnostopoulou, 2022: Reliability of the ERA5 in replicating mean and extreme temperatures across Europe. *Water*, **14** (4), URL <https://www.mdpi.com/2073-4441/14/4/543>.
- Vitart, F., A. W. Robertson, and D. L. Anderson, 2012: Subseasonal to seasonal prediction project: Bridging the gap between weather and climate. *Bulletin of the World Meteorological Organization*, **61** (2), 23.
- Vitart, F., and Coauthors, 2008: The new VAREPS-monthly forecasting system: A first step towards seamless prediction. *Quart. J. Roy. Meteor. Soc.*, **134** (636), 1789–1799.
- Vitart, F., and Coauthors, 2017: The subseasonal to seasonal (s2s) prediction project database. *Bull. Amer. Meteor. Soc.*, **98** (1), 163 – 173, <https://doi.org/10.1175/BAMS-D-16-0017.1>.
- Vitart, F., and Coauthors, 2019: Extended-range prediction. Tech. Rep. 854, ECMWF. URL <https://www.ecmwf.int/node/19286>.
- von Storch, H., F. Zwiers, and C. U. Press, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, URL [https://books.google.fr/books?id=\\_VHxE26QvXgC](https://books.google.fr/books?id=_VHxE26QvXgC).
- Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the northern hemisphere winter. *Mon. Wea. Rev.*, **109** (4), 784 – 812, [https://doi.org/10.1175/1520-0493\(1981\)109<0784:TITGHF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2).
- Wang, X. L., and F. W. Zwiers, 2001: Using redundancy analysis to improve dynamical seasonal mean 500 hPa geopotential forecasts. *Int. J. Climatol.*, **21** (5), 637–654, <https://doi.org/10.1002/joc.638>.
- White, C. J., and Coauthors, 2017: Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteor. Appl.*, **24** (3), 315–325, <https://doi.org/10.1002/met.1654>.
- Wilby, R. L., and T. M. Wigley, 1997: Downscaling general circulation model output: a review of methods and limitations. *Progress in physical geography*, **21** (4), 530–548.
- Wilcoxon, F., 1945: Individual comparisons by ranking methods. *Biometrics Bulletin*, **1** (6), 80–83, URL <http://www.jstor.org/stable/3001968>.
- Wilks, D. S., 2013: Projecting “Normals” in a Nonstationary Climate. *J. Appl. Meteor. Climatol.*, **52** (2), 289–302, <https://doi.org/10.1175/JAMC-D-11-0267.1>.
- Wilks, D. S., 2014: Comparison of Probabilistic Statistical Forecast and Trend Adjustment Methods for North American Seasonal Temperatures. *J. Appl. Meteor. Climatol.*, **53** (4), 935–949, <https://doi.org/10.1175/JAMC-D-13-0294.1>.

- Wilks, D. S., 2019: *Statistical methods in the atmospheric sciences*. 4th ed., Elsevier, Cambridge.
- Wilks, D. S., and R. E. Livezey, 2013: Performance of Alternative “Normals” for Tracking Climate Changes, Using Homogenized and Nonhomogenized Seasonal U.S. Surface Temperatures. *J. Appl. Meteor. Climatol.*, **52** (8), 1677–1687, <https://doi.org/10.1175/JAMC-D-13-026.1>.
- WindEurope, 2022: Wind energy in Europe: 2021 statistics and the outlook for 2022-2026. Tech. rep., Rue Belliard 40, B-1040 Brussels. URL <https://tinyurl.com/5bydexpd>.
- Woolnough, S. J., F. Vitart, and M. A. Balmaseda, 2007: The role of the ocean in the Madden–Julian Oscillation: Implications for MJO prediction. *Quart. J. Roy. Meteor. Soc.*, **133** (622), 117–128, <https://doi.org/10.1002/qj.4>.
- Zhang, F., Y. Q. Sun, L. Magnusson, R. Buizza, S.-J. Lin, J.-H. Chen, and K. Emanuel, 2019a: What is the predictability limit of midlatitude weather? *J. Atmos. Sci.*, **76** (4), 1077 – 1091, <https://doi.org/10.1175/JAS-D-18-0269.1>.
- Zhang, F., Y. Q. Sun, L. Magnusson, R. Buizza, S.-J. Lin, J.-H. Chen, and K. Emanuel, 2019b: What is the predictability limit of midlatitude weather? *J. Atmos. Sci.*, **76** (4), 1077 – 1091, <https://doi.org/10.1175/JAS-D-18-0269.1>.
- Zheng, C., E. K.-M. Chang, H.-M. Kim, M. Zhang, and W. Wang, 2018: Impacts of the Madden–Julian Oscillation on storm-track activity, surface air temperature, and precipitation over North America. *J. Climate*, **31** (15), 6113 – 6134, <https://doi.org/10.1175/JCLI-D-17-0534.1>.
- Zhu, H., M. C. Wheeler, A. H. Sobel, and D. Hudson, 2014: Seamless Precipitation Prediction Skill in the Tropics and Extratropics from a Global Model. *Mon. Wea. Rev.*, **142** (4), 1556–1569, <https://doi.org/10.1175/MWR-D-13-00222.1>, publisher: Amer. Meteor. Soc. Section: Mon. Wea. Rev.
- Zorita, E., and H. von Storch, 1999: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *J. Climate*, **12** (8), 2474 – 2489, [https://doi.org/10.1175/1520-0442\(1999\)012<2474:TAMAAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2).
- Žagar, N., and I. Szunyogh, 2020: Comments on “what is the predictability limit of midlatitude weather?”. *J. Atmos. Sci.*, **77** (2), 781 – 785, <https://doi.org/10.1175/JAS-D-19-0166.1>.

## Supplemental material

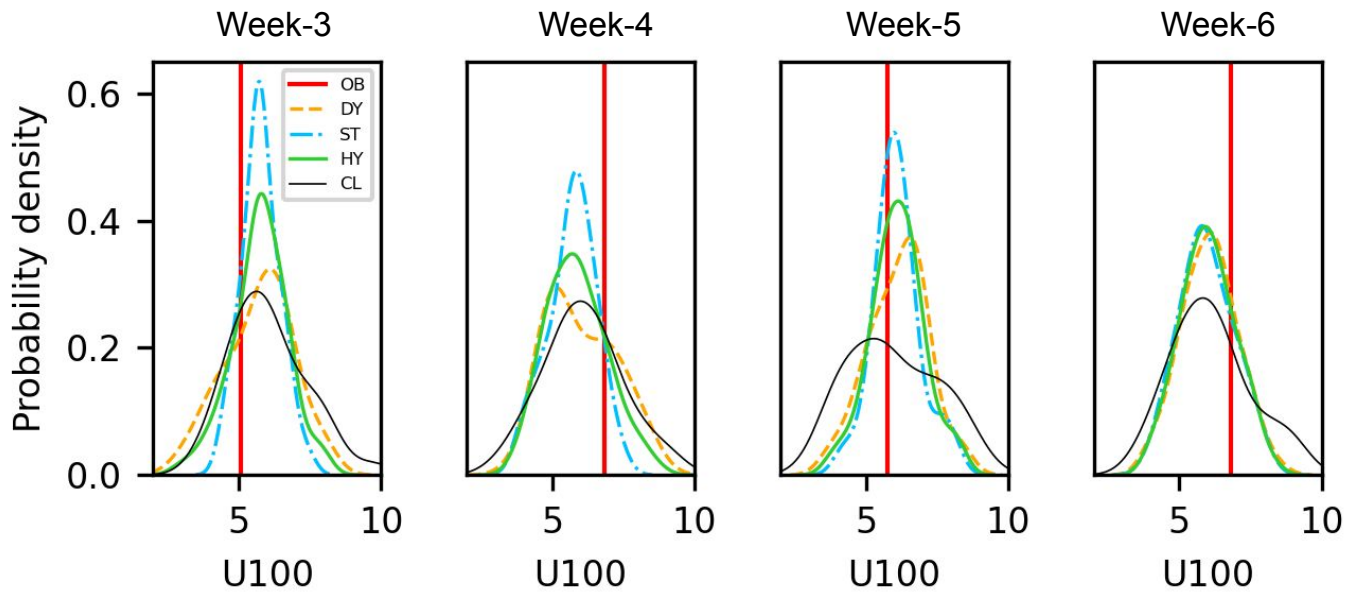


**Figure S1:** Illustration of the temporal evolution of ensemble Probability Density Functions (PDFs) of dynamical (DY), statistical (ST), hybrid (HY), and climatological (CL) 100 m wind speed predictions. The PDFs are computed as kernel-density estimates (Gaussian kernel) using ensemble members of weekly mean values averaged over **Germany** ( $47.3^{\circ}$  -  $55^{\circ}$  N,  $6.3^{\circ}$  -  $15.4^{\circ}$  E). The grid point values are weighted by the cosine of their respective latitude before computing domain average. This illustration corresponds to dynamical predictions initiated on 5 February 2018. The red vertical line in each of the panels indicates the observed weekly mean (OB).

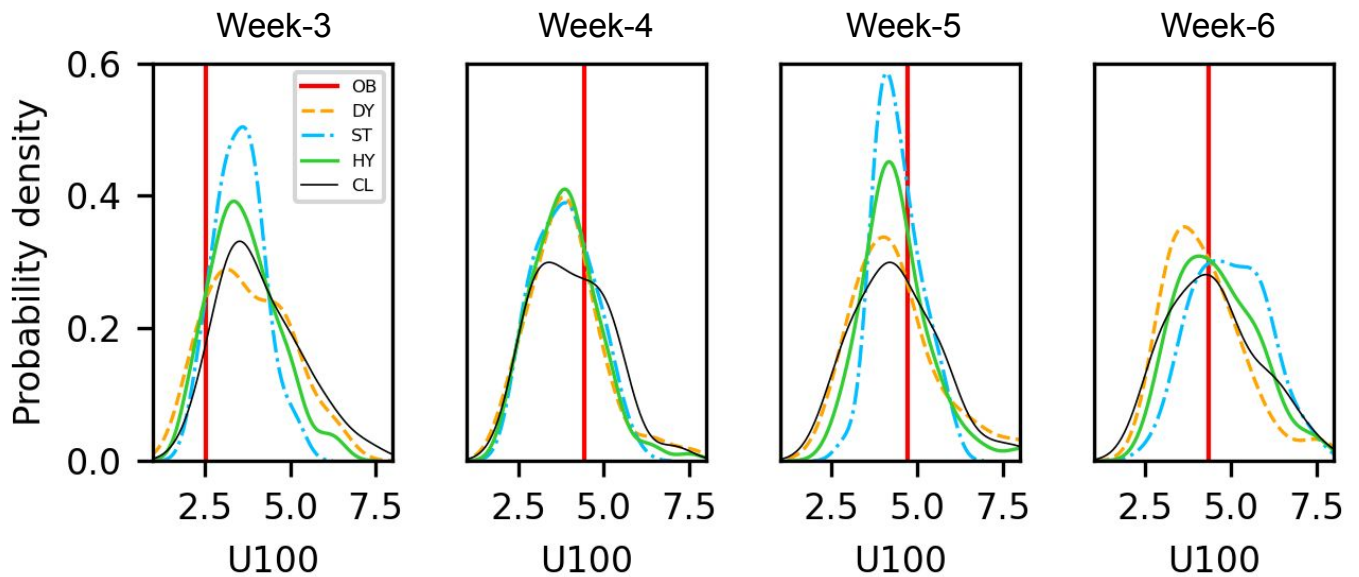


**Figure S2:** As in Figure S1, but over the **United Kingdom** ( $49^{\circ}$  -  $60^{\circ}$  N,  $10^{\circ}$  W $^{\circ}$  -  $4^{\circ}$  E) corresponding to dynamical predictions initiated on 5 January 2017.

Supplemental material



**Figure S3:** As in Figure S1, but over **France** ( $43^{\circ}$  -  $51.4^{\circ}$  N,  $5.5^{\circ}$  W $^{\circ}$  -  $7.3^{\circ}$  E) corresponding to dynamical predictions initiated on 3 January 2019.



**Figure S4:** As in Figure S1, but over **Spain** ( $36^{\circ}$  -  $43.3^{\circ}$  N,  $7.2^{\circ}$  W $^{\circ}$  -  $0.1^{\circ}$  E) corresponding to dynamical predictions initiated on 23 December 2019.





# 4 Case study: How well in advance can we anticipate wind droughts?

Under pressure, we do not rise to the occasion,  
but default to our level of training.  
— Anonymous

## Objective

The key objective of this chapter is to understand the limits of operational predictability of the European wind droughts through the case of the July 2018 wind drought.

## Data and Methods

As forecasts, we use the sub-seasonal predictions of 100-m wind speed over Europe and geopotential height at 500 hPa over Euro-Atlantic from the European Centre for Medium-Range Weather Forecasts for boreal summer between 2016 and 2021. As a reference, we use ERA5 reanalysis of the same two variables over the same domain for boreal summer between 1979 and 2021. We obtain hybrid predictions using the statistical downscaling methodology developed in the previous chapter. We then investigate the predictability limits of both dynamical and hybrid predictions in anticipating the wind drought of July 2018.

## Key conclusion

Both the dynamical and hybrid predictions fail to anticipate the wind drought beyond a lead time of one week. This is attributed to persistent difficulties of numerical weather prediction models with atmospheric blocking.

## Publication

The manuscript of this chapter has been submitted to Monthly Weather Review in October 2022, and is currently under review.

This manuscript has been submitted to Monthly Weather Review in October 2022, and is currently under review.

## On the sub-seasonal predictability of the European wind droughts: a case of July 2018

NAVEEN GOUTHAM,<sup>a,b</sup> RIWAL PLOUGONVEN,<sup>b</sup> HIBA OMRANI,<sup>a</sup> SYLVIE PAREY,<sup>a</sup> AND PHILIPPE DROBINSKI<sup>b</sup>

<sup>a</sup> EDF Lab Paris-Saclay, Palaiseau, France.

<sup>b</sup> Laboratoire de Météorologie Dynamique-IPSL, Ecole Polytechnique, Institut Polytechnique de Paris, ENS, PSL Research University, Sorbonne Université, CNRS, France.

**ABSTRACT:** The European energy sector is increasingly becoming weather-sensitive due to the growing share of variable renewables in the electricity mix. With Europe aiming to derive 40% of its demand from renewables by 2030, several recent episodes of wind droughts have raised concerns about the security of supply. Hence, the energy sector needs to anticipate such extreme cases to plan for alternatives. With this motivation, we study the sub-seasonal predictability of the July 2018 wind drought over Europe when 100-m wind speeds weaker than 30% of the climatological mean were observed for more than a week over the North Sea. These weak winds result from blocked circulation over eastern Scandinavia and the Atlantic Ocean. Assessment of sub-seasonal predictions from the European Centre for Medium-Range Weather Forecasts (ECMWF) indicates high uncertainty in predicting both the upper-level as well as surface fields a week before the event. Additionally, this study allowed us to test the application of a recently developed statistical downscaling methodology in improving the sub-seasonal wind speed predictions. No significant improvements are obtained relative to dynamical predictions, consistent with the difficulty in predicting large-scale atmospheric evolution. Our analysis attributes the failure of predicting this wind drought to the persistent difficulties of numerical weather prediction models with atmospheric blocking. Finally, we investigate the skill of the ECMWF model in predicting extreme winds by comparing the Brier Skill Score between upper- and lower-decile predictions. The results suggest a bias of the ECMWF model in more skillfully predicting one extreme compared with the other.

### 1. Introduction

The European Green Deal, approved in 2021, will make Europe the world's first climate-neutral continent by 2050 if implemented correctly (IEA 2020). This deal comes in the backdrop of the successful implementation of the earlier 20/20/20 strategy of the European Union (Förster et al. 2021). To achieve climate neutrality by 2050, the European Commission has set an ambitious interim target of deriving 40% of its final energy demand from renewable sources by 2030 (IEA 2021). This target of 40% is almost double the share met by renewable sources in 2019.

In 2019, wind power became the largest renewable source of electricity in the European Union (IEA 2020). With a growing share of variable renewable power systems in the energy mix, the European energy sector is increasingly becoming weather-sensitive and thus concerned about the security of supply. Several recent episodes of wind droughts (i.e., prolonged periods with weak winds) and their associated impacts on the energy sector are a testimony to this (e.g., TheGuardian 2018; NewScientist 2018). One such episode occurred in July 2018 when wind speeds weaker than 30% of the climatological normal were observed for more than a week in and around the North Sea. Figure 1 illustrates the 100-m wind speed for the week between 16 and 22 July 2018. As can be noticed from the observed weekly mean wind speed (left panel in Figure 1), the wind speed in and around the North sea is typically lower or closer to the cut-in speed (i.e., between

3 and 4 m s<sup>-1</sup> (Manwell et al. 2010; Burton et al. 2011)) of a wind turbine. The right panel in Figure 1 shows pronounced negative anomalies (down to -50%) in and around the North Sea, with the average being -26% in the region bounded by the magenta-colored rectangle. A 26% lower wind speed translates to about 60% reduction in wind energy production relative to climatology. The North Sea, with the highest density of installed wind power capacity in Europe at the time of writing this manuscript (Figure 2) (WindEurope 2022), is a crucial contributor to the European energy mix. Despite the electricity demand being lower and the solar photovoltaic production being higher in summer compared with other seasons (e.g., Šúri et al. 2007; Gallo Cassarino et al. 2018), such low wind situations may put the energy sector at serious risk of undersupply. In this regard, a few studies have investigated the episodes of wind droughts in the past: Jiménez et al. (2011) studied the effect of heat waves and drought on surface wind conditions during the summer of 2003 over the northeastern Iberian Peninsula and found that the wind drought was due to synoptic conditions blocking the climatological north-westerly flow. Additionally, they linked the wind drought to the modulation of mesoscale conditions influenced by the availability of soil moisture. Lledó et al. (2018) examined the effect of Pacific sea surface temperature on the wind drought which affected the United States in the first quarter of 2015. They found high sea surface temperatures over the western tropical Pacific linked to a strongly positive phase of the Northern Pacific Mode to be the main driver of the wind drought. Ohlendorf and Schill (2020)

---

Corresponding author: Naveen Goutham, naveen.goutham@edf.fr/naveen.goutham@polytechnique.edu

investigated the frequency and duration of low-wind events for onshore wind power in Germany using 40 years of re-analysis data. They found that the frequency of low-wind events is higher in summer than in winter. Li et al. (2021) studied the climatology of coupled low wind and solar power events (a.k.a. *Dunkelflaute* events) in 11 countries surrounding the North and the Baltic seas. They found an increased likelihood of the occurrence of the *Dunkelflaute* events in late fall and early winter compared with the other months of the year. Nevertheless, based on the knowledge of the authors, there is no published peer-reviewed research on the operational predictability of wind droughts.

In recent times, there is a growing interest within the energy sector about sub-seasonal predictions, i.e. predictions beyond two weeks and up to two months (Robertson and Vitart 2018). The predictability on sub-seasonal timescales is primarily provided by the Madden-Julian oscillation (e.g., Jones et al. 2004a,b; Zheng et al. 2018), snow cover (e.g., Sobolowski et al. 2010; Lin and Wu 2011; Orsolini et al. 2013), stratosphere-troposphere interactions (e.g., Baldwin et al. 2003; Domeisen et al. 2020; Schwartz and Garfinkel 2020), land (e.g., Koster et al. 2011; van den Hurk et al. 2012; Prodhomme et al. 2016; Seo et al. 2019), and ocean conditions (e.g., Woolnough et al. 2007; Fu et al. 2007; Subramanian et al. 2019). Having prior information about the anticipated renewable energy production and electricity consumption on sub-seasonal timescales can guide the energy industry in planning the required reserve levels, scheduling asset maintenance, determining grid transmission capacity, and trading energy in markets (White et al. 2017; Soret et al. 2019). To support the future growth of wind power in Europe, it is important to understand the operational predictability and the limits thereof of wind droughts. Hence, we focus this study on the sub-seasonal predictability of the July 2018 wind drought over Europe. The manuscript is organized as follows: section 2 describes the data and methodology employed, section 3 analyzes the predictability of the July 2018 wind drought in detail, and section 4 provides discussion and conclusions.

## 2. Data and Methodology

### a. Data

In this study, we use extended-range forecasts and retrospective forecasts (re-forecasts) originating from the European Centre for Medium-Range Weather Forecasts (ECMWF) (Vitart et al. 2019). The extended-range forecasts are produced by extending the ocean-atmosphere coupled medium-range ensemble forecasts (i.e., up to two weeks) to 46 days twice a week at 00 UTC on Mondays and Thursdays. These forecasts are run at a spatial resolution of  $\sim 18$  km (Tco639L91) up to a lead time of 15 days, and at  $\sim 36$  km (Tco319L91) beyond (Vitart et al. 2017, 2019). The operational ensemble forecasts consist of 51 members (50 perturbed + control), of which the perturbed members

are obtained by perturbing both the initial conditions and the model physics. The perturbations in the initial conditions are obtained using singular vectors (Leutbecher 2005; Leutbecher and Palmer 2008) and ensemble data assimilation (Buizza et al. 2008; Isaksen et al. 2010). The model perturbations are carried out using the Stochastically Perturbed Parametrisation Tendencies (SPPT) scheme (Buizza et al. 1999; Palmer et al. 2009; Leutbecher et al. 2016). The perturbed members thus obtained serve to account for uncertainties in initial conditions and model parametrizations (Buizza 2019).

Despite accounting for uncertainties in initial conditions and model parametrizations, the forecasts suffer from systematic errors (Siegert and Stephenson 2019). The systematic errors are generally in the form of persistent over- or under-estimation of values and over- or under-dispersion of the distribution. It is necessary to remove systematic errors before employing the model. The ECMWF produces re-forecasts which serve to estimate and remove systematic errors inherent in the operational version (Vitart et al. 2008). For each issued operational forecast, the ECMWF produces a set of 20 re-forecasts of 11 members (10 perturbed + control), issued for the same calendar day of the year as the operational forecast over each of the past 20 years. These re-forecasts are initialized using initial conditions obtained from ERA5 reanalysis. All the forecasts employed in this study are bias corrected using the Mean and Variance Adjustment method (Torralba et al. 2017; Manzananas et al. 2019; Goutham et al. 2022).

We retrieve forecasts and the corresponding re-forecasts of zonal and meridional components of 100-m wind speed and geopotential at 500 hPa on a global grid from the Meteorological Archival and Retrieval System of the ECMWF. The retrieved spatial resolution is  $0.9^\circ$  and the temporal resolution is 6 h (instantaneous values at 00, 06, 12, and 18 UTC). The data are retrieved for the boreal summer (i.e., June-July-August) between 2016 and 2021. We obtain 100 m wind speed by computing the square root of the sum of squares of the zonal and meridional components. The geopotential height ( $Z_{500}$ ) is computed by dividing the geopotential by the Earth's gravitational acceleration,  $g$  ( $= 9.806 \text{ m s}^{-2}$ ). We use only the perturbed members of forecasts and re-forecasts in this study. The control of the ensemble should be treated as another indistinguishable ensemble member. However, due to the unavailability of the control member in the internal database of ESPRI (Ensemble de services pour la recherche à l'Institut Pierre-Simon Laplace) as a result of an unintentional man-made error, we had to use only the perturbed members. Nevertheless, we expect the impact of the missing control on the properties of the ensemble distribution on sub-seasonal timescales to be insignificant.

Although the forecast quality is generally verified against observations (Coelho et al. 2019; Wilks 2019), we use ERA5 reanalysis (Hersbach et al. 2020) as the ground

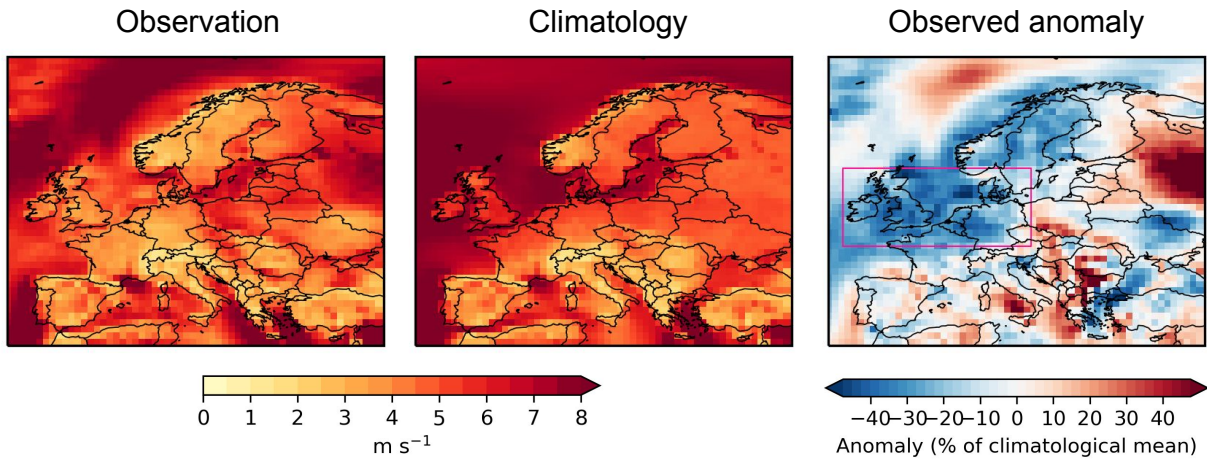


FIG. 1. Maps of the 100-m weekly mean wind speed for the week between 16 and 22 July 2018, the mean climatology for July (1979-2021 without 2018), and the corresponding anomaly over Europe. The figures are produced using data from ERA5 reanalysis (refer to section 2a)

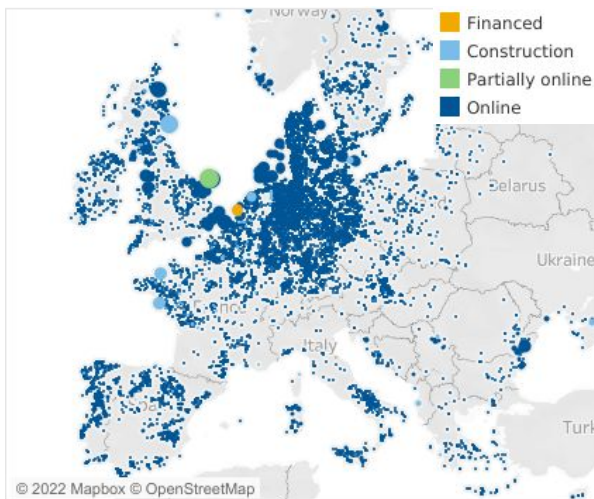


FIG. 2. Location of the major onshore and offshore wind farms in Europe as of August 2022. The map is sourced from Wind Europe (<https://windeurope.org/>).

truth due to the limited- or non-availability of serially complete and spatially coherent observed dataset (Kalnay 2003; Coelho et al. 2019). We retrieve the same variables as stated earlier from the Climate Data Store of the Copernicus Climate Change Services (Raoult et al. 2017) at the same temporal and spatial resolution as the forecasts. The data are retrieved for the period between 1979 and 2021. The choice of using ERA5 reanalysis as the ground truth is influenced by the fact that the wind speed is represented more realistically in ERA5 reanalysis when compared with other reanalysis datasets over Europe (e.g., Olauson 2018; Ramon et al. 2019; Jourdi er 2020; D orenk amper et al.

2020; Brune et al. 2021; Molina et al. 2021; Murcia et al. 2022).

#### b. Methodology

Despite identifying the fundamental sources of predictability on sub-seasonal timescales (Vitart et al. 2012), the skill horizon of sub-seasonal predictions of 100-m wind speed is between two to three weeks over Europe, depending on the season and geographic region (Goutham et al. 2022). In addition to the increased sensitivity of surface variables to model parametrizations (Buizza and Leutbecher 2015; Toth and Buizza 2019), the difficulties of numerical weather prediction models in accurately forecasting surface variables beyond two weeks are linked to our improper understanding of the complex interactions between various components of the Earth system (Palmer et al. 2009; Leutbecher et al. 2016; Robertson and Vitart 2018; Lled o and Doblas-Reyes 2020). Nevertheless, taking into consideration the longer skill horizon of large-scale, upper-level fields compared with surface fields, and the physical relationship between them, we can employ statistical downscaling techniques to infer surface variables using the more reliable information contained in the prediction of large-scale fields (e.g. Alonzo et al. 2017; Alonzo 2018; Manzanas et al. 2018; Lled o and Doblas-Reyes 2020; Goutham et al. 2021).

In this study, we implement a statistical downscaling methodology which we have developed in Goutham et al. (Forthcoming). Figure 3 summarizes the methodology implemented in this study. Overall, the statistical downscaling is carried out in two steps:

1. **Capturing the relationship between the large-scale atmospheric circulation and 100-m wind speed:** Here, we employ redundancy analysis (von Storch

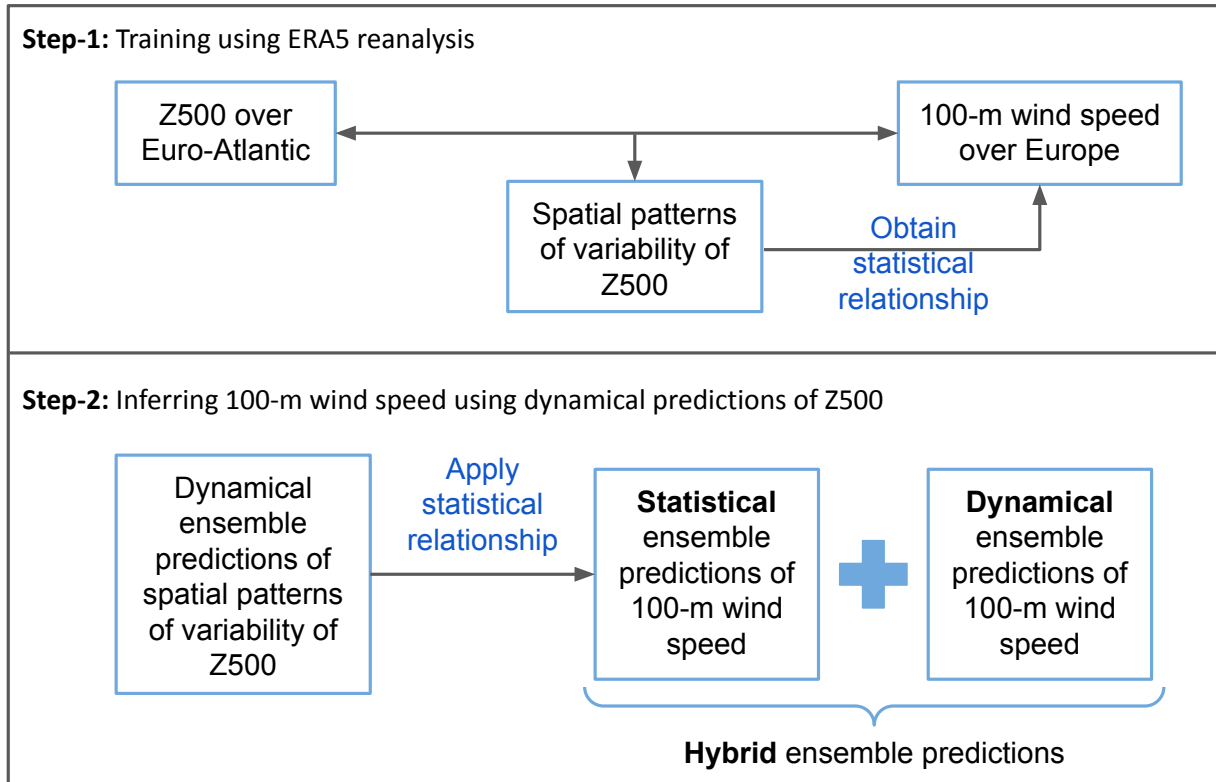


FIG. 3. Illustration of the statistical downscaling methodology employed in this study.

et al. 1999; Tippett et al. 2008; Wilks 2014, 2019) between geopotential height at 500 hPa over the Euro-Atlantic and 100-m wind speed over Europe to obtain spatial patterns of variability of Z500 that maximize the explained variance of 100-m wind speed. The training in this study is carried out using ERA5 reanalysis for boreal summer between 1999 and 2015.

2. **Application on sub-seasonal predictions of the large-scale atmospheric circulation:** We apply the linear relationship obtained in the previous step on the sub-seasonal ensemble predictions of a restricted number of spatial patterns of variability of Z500 to infer 100-m wind speed over Europe. We call the reconstructed ensemble predictions of 100-m wind speed *statistical predictions*. We then combine these predictions with dynamical ensemble predictions of 100-m wind speed to obtain *hybrid predictions* having twice the size as dynamical predictions. In this study, the redundancy analysis model is employed on sub-seasonal predictions initialized during boreal summer between 2016 and 2021.

We refer the reader to Goutham et al. (Forthcoming) for a detailed description and the mathematical formulation of the methodology.

### 3. Analysis of the case

#### a. The cause

The surface weather over Europe is mainly driven by the large-scale atmospheric circulation over the Euro-Atlantic (Grams et al. 2017; Zubiate et al. 2017; van der Wiel et al. 2019; Bloomfield et al. 2019; Cortesi et al. 2019; Garrido-Perez et al. 2020; Cortesi et al. 2021). Figure 4 illustrates the geopotential height at 500 hPa (Z500), the climatology, and the anomaly observed for the week between 16 and 22 July 2018. The anomaly figure (right panel) shows two positive anomalies around Europe: one over eastern Scandinavia, and the other over the Atlantic Ocean. The geopotential height over eastern Scandinavia is about 180 m above the climatological normal, and that over the Atlantic Ocean is about 140 m above the climatological normal. Such blocking patterns displace the polar jet more northward, thereby blocking the typical westerly flow from entering Europe (e.g., Rex 1950; Reinhold 1987; Vautard 1990; Michelangeli et al. 1995). This results in weak winds in and around the North Sea (e.g., Grams et al. 2017; Zubiate et al. 2017; van der Wiel et al. 2019; Bloomfield et al. 2019; Cortesi et al. 2019; Garrido-Perez et al. 2020; Cortesi et al. 2021).



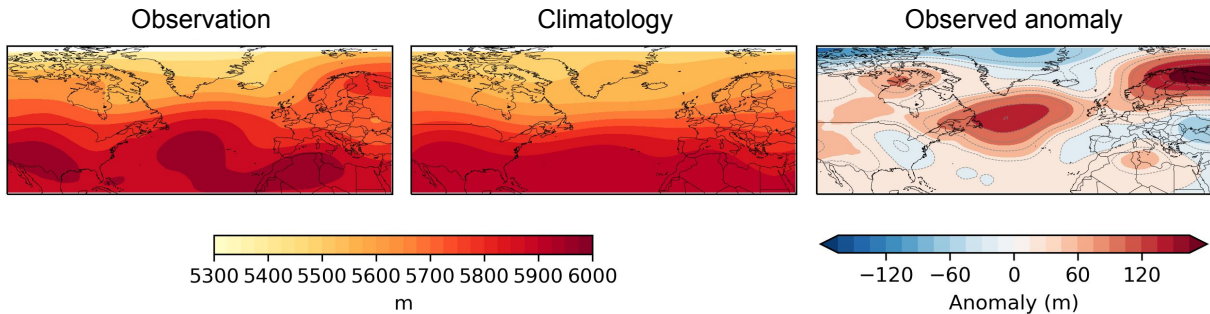


FIG. 4. As in Figure 1, but for Z500. The dashed lines in the right panel indicate isobars and are plotted at an interval of 30 m.

#### b. How well was the wind drought forecast?

Figure 5 shows the weekly mean 100-m wind speed anomalies for the week between 16 and 22 July 2018 from the forecasts initialized on 16 July (lead-W0) and 9 July (lead-W1). In lead-W0, almost all the members of the ensemble indicate negative anomalies over the North Sea and positive anomalies over western Russia, demonstrating a similar intensity and spatial pattern to that of the observed anomaly (right panel in Figure 1). However, in lead-W1, the ensemble displays a large spread, with members showing contrasting anomalies. Although *five out of fifty* members predict both the spatial pattern and the intensity closer to the observed anomaly (not shown), the ensemble mean neither predicts the intensity nor the spatial pattern right. To illustrate this last point, the ensemble mean over-predicts wind speed over the United Kingdom on average by about 15%, and under-predicts over western Russia by about 50%.

To understand the information contained in the ensemble, we illustrate in Figure 6 the weekly evolution of accuracy of the dynamical (DY) and hybrid (HY) ensemble predictions, as a function of lead time, in forecasting weak winds for the week between 16 and 22 July 2018. For this purpose, we consider forecasts initialized on several dates: 16 July (Monday), 9 July (Monday), and 11 June 2018 (Monday). The accuracy is measured as the probability of observing wind speed within  $1 \text{ m s}^{-1}$  around the observation. Accordingly, the higher the probability, the higher the accuracy. An allowance of  $1 \text{ m s}^{-1}$  of the observation, although arbitrary, is chosen to accommodate for the model errors.

In lead-W0 in Figure 6, we can observe high accuracy of the ensemble dynamical forecast over a large part of Europe (average probability of 80% over the European domain), higher than that of climatology (European domain average of 47%). Kindly note that the hybrid prediction is not produced for lead-W0 as the forecast of Z500 is deterministic in this lead. For more information, we refer the reader to Goutham et al. (Forthcoming). In lead-W1, the over-

all accuracy of dynamical prediction drops significantly to about 51% (European domain average). The hybrid prediction, with a European domain-averaged probability of 55% in lead-W1, is more accurate than the dynamical prediction. The European domain-averaged probability of statistical prediction alone in lead-W1 (59%, not shown) is higher than that of both dynamical and hybrid predictions. Since the probability of obtaining wind speed within  $1 \text{ m s}^{-1}$  around the observation in all the three predictions saturates starting lead-W1, we skip lead-W2, lead-W3, and lead-W4 in the figure. In lead-W5, the European domain-averaged probabilities of dynamical and hybrid predictions are 47% and 51%, respectively. In lead-W1 and prior, both dynamical and hybrid predictions completely miss both regions of strong anomalies: the weak winds over the North Sea and the strong winds over western Russia (probabilities less than 20%). Nevertheless, hybrid predictions generally demonstrate higher accuracy compared with dynamical predictions in predicting *normal* winds, i.e. winds close to their climatological value. This can be appreciated by comparing the spatial structure and the corresponding intensity of blue regions in hybrid predictions with the normal winds in the observed anomaly map (i.e., white regions). Overall, none of the forecasts are able to anticipate weak and strong winds beyond a week with an acceptable level of certainty.

#### c. How well was the synoptic state forecast?

In Goutham et al. (Forthcoming), we illustrated that the accuracy of 100-m wind speed hybrid predictions is generally higher than that of dynamical predictions over a large part of Europe. In addition, we also demonstrated the longer skill horizon of hybrid predictions compared with dynamical counterparts. Despite this, the 100-m wind speed hybrid predictions failed to foresee weak winds beyond a week in the case of the July 2018 wind drought. The failure of the redundancy analysis model, in this case, may be attributed to one or both of the following:

1. *The breaking of the key hypothesis*

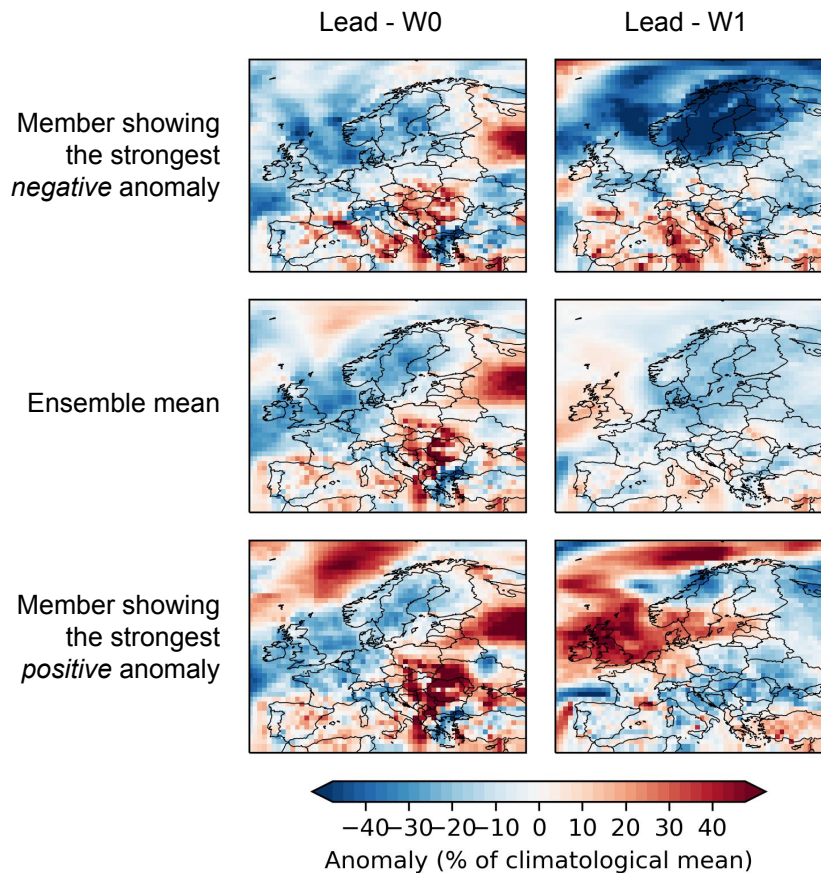


FIG. 5. Illustration of 100-m wind speed anomaly fields, showing extreme members of the ensemble and the ensemble mean for the week between 16 and 22 July 2018. Lead-W0 corresponds to the forecasts initialized on 16 July, and Lead-W1 corresponds to those initialized on 9 July.

To recall, the key hypothesis of the redundancy analysis model is that the predictions of large-scale, upper-level fields are generally more reliable than those of surface fields (section 2b). Does this hypothesis hold during extreme events such as wind droughts?

2. *Insufficient learning by the model due to the rarity of the event*

The large-scale atmospheric state which led to this wind drought may be rare and specific, and hence the model may not have learned to treat such a case during the training phase.

We first address the question of the key hypothesis of the model. To understand the reliability of the prediction of the large-scale atmospheric state, we illustrate in Figure 7 the state of weekly mean Z500 for the week between 16 and 22 July 2018 as it was forecast in the dynamical predictions initialized on 16 July (lead-W0) and 9 July 2018 (lead-W1). Similar to Figure 5, almost all the ensemble members predict blocking over eastern Scandinavia and

the Atlantic Ocean in lead-W0. However, in lead-W1, the ensemble displays a large spread. A large proportion of the ensemble members either miss the blocking over eastern Scandinavia and the Atlantic Ocean completely or underestimate the intensity starting lead-W1. In addition to wrongly predicting a depression over the Atlantic Ocean in lead-W1, the ensemble mean underestimates the eastern Scandinavian blocking in excess of 100 m.

To further understand the ability of the ECMWF model in forecasting atmospheric blocking over Europe that led to the wind drought in and around the North Sea, we summarize blocking by computing the indices of GHGN (Geopotential Height gradients North) and GHGS (Geopotential Height gradients South) over Europe (LejenÅs and Økland 1983; Tibaldi and Molteni 1990) (Appendix a). Figure 8 illustrates the daily evolution of blocking indices. Overall, from Figure 8, it is conspicuous that the model finds it difficult to predict blocking or anticipate its temporal evolution precisely (when initialized after the blocking onset) after about a week. This can be illustrated by tracking the



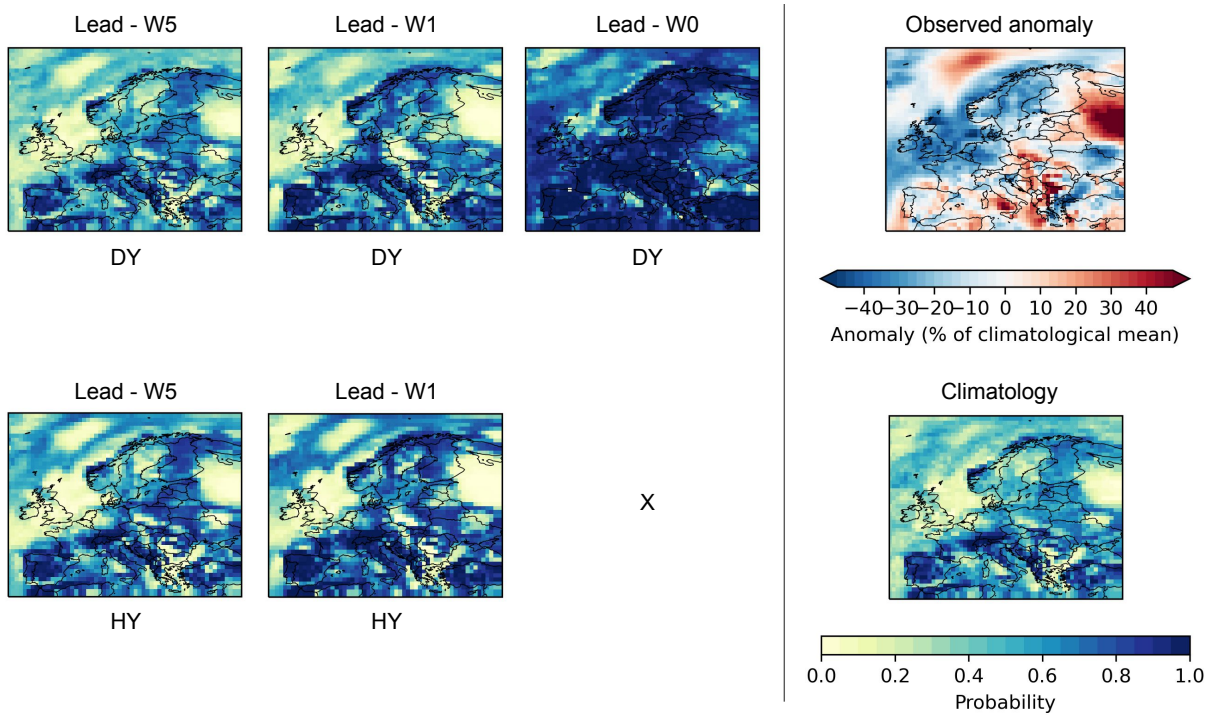


FIG. 6. Illustration of the weekly evolution of accuracy of the different ensemble forecasts, as a function of lead time, for the week between 16 and 22 July 2018, showing the probability of obtaining the 100-m wind speed within  $1 \text{ m s}^{-1}$  around the observation. The observed anomaly for the week is shown in the top right panel. The climatological probabilities are shown in the bottom right panel. The first three columns from left to right correspond to forecasts initialized on 11 June, 9 July, and 16 July, respectively. DY = Dynamical predictions and HY = Hybrid predictions.

temporal evolution of the blue and orange curves from the forecasts initialized on 9 July 2018. Despite all three forecasts being initialized with a blocked state with  $\text{GHGN} < -10 \text{ m}^{\circ}\text{lat}^{-1}$ , the forecasts find it difficult to follow the temporal evolution of blocking after about six days. Although the GHGS from the forecast initialized on 9 July 2018 is well correlated in time with that of the observation, the forecast grossly underestimates the observed GHGS after about a week. As already pointed out in the literature, this case highlights the persistent difficulties of numerical weather prediction models in forecasting the onset and decay of atmospheric blocking events (e.g., Tibaldi and Molteni 1990; Ferranti et al. 2015; Matsueda and Palmer 2018; Büeler et al. 2021; Lupo 2021; Davini et al. 2021; Kleiner et al. 2021; Cortesi et al. 2021).

We now address the other possible reason for the failure of hybrid prediction in anticipating the July 2018 wind drought. We particularly try to understand whether the statistical downscaling model employed in this study has learned sufficiently to reconstruct 100-m wind speed over Europe using Z500 over Euro-Atlantic as the predictor. For this purpose, we test the statistical downscaling model using the observed weekly mean anomaly of Z500 for the week between 16 and 22 July 2018 as the predictor (right

panel in Figure 4). Figure 9 shows the corresponding predicted 100-m wind speed anomaly over Europe (to be compared to the right panel of Figure 1). Overall, despite the ability of the statistical downscaling model in predicting weak winds over the United Kingdom, Germany, and the region around the North Sea, the model fails to reproduce the intensity right ( $-15\%$  over the North Sea instead of  $-30\%$ ). Furthermore, the reconstructed wind speed anomaly does not reproduce the spatial pattern, especially over northern Europe. The positive wind anomaly in western Russia is also absent. This suggests that the model is insufficiently trained on extreme cases such as the one presented here. This can be elucidated in Figure 10 which shows the climatology of blocking indices for the boreal summer between 1999 and 2015, i.e. the period used for training the statistical downscaling model. The weekly mean indices of GHGN and GHGS for the week between 16 and 22 July 2018 are illustrated in the figure as black solid lines. With GHGN being closer to the 10th percentile and GHGS being greater than the 95th percentile, it is conspicuous that the atmospheric blocking of 16-22 July 2018 is a rare event. In addition to insufficient learning by the statistical downscaling model, the substandard performance of the model in reconstructing surface

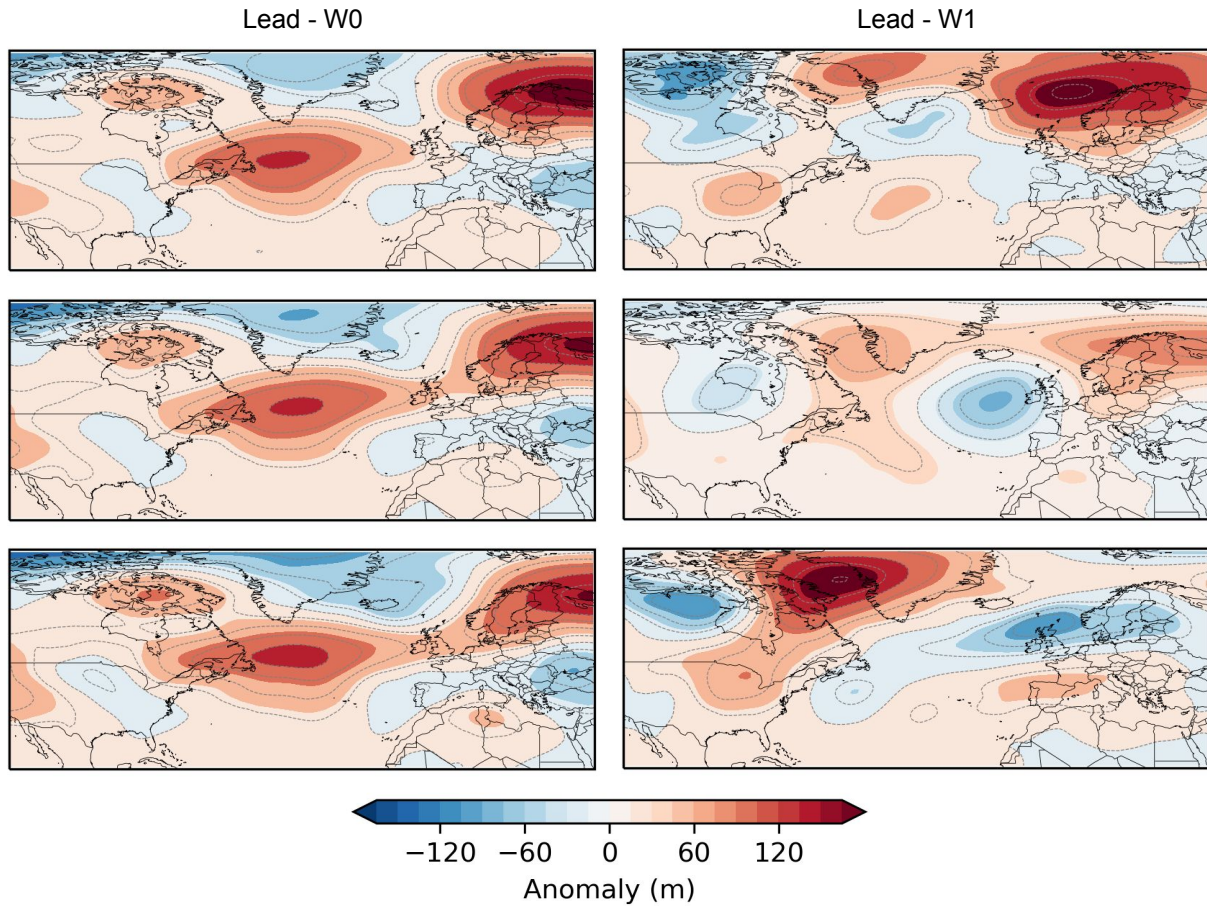


FIG. 7. As in Figure 5, but for Z500 anomalies. The members illustrated here correspond to those shown in Figure 5.

winds may be attributed to the inadequate statistical predictive power of a sole predictor (Z500). The way forward to improving the predictive capability of the redundancy analysis model is by retraining the model using data covering a longer period than what is currently being used, for instance, using the back-extended ERA5 reanalysis (Bell et al. 2021). Furthermore, additional predictors such as sea surface temperature, geopotential height at the level of the stratosphere, and land conditions can be used to better predict atmospheric blocking (e.g., Jiménez et al. 2011; Miller and Wang 2019), thereby improving the prediction of 100-m wind speed.

#### 4. Discussion and Conclusions

The previous section has highlighted the difficulties of the dynamical and hybrid predictions in anticipating extremely weak winds over Europe for the July 2018 case. This raises a broader question on the skill of dynamical and hybrid prediction models to forecast *extreme* winds, weak

or strong, in general. To investigate this, we compute the Brier Skill Score (BSS) (Brier 1950; Jolliffe and Stephenson 2003; Coelho et al. 2019; Wilks 2019) for the upper and lower decile of 100-m wind speed, in dynamical and hybrid predictions over Europe. The Brier Skill Score measures the relative accuracy of a probabilistic forecasting system with respect to another in forecasting a dichotomous event (i.e., yes or no event). The Brier Skill Score is based on the Brier Score, which is essentially the mean squared error between forecast probabilities and the corresponding binary observations. The Brier Score is negatively oriented. Accordingly, a perfect forecasting system should have a Brier Score of zero. The reader is referred to Appendix b to understand the computation of the Brier Skill Score. Overall, the skill horizon of both dynamical and hybrid predictions in foreseeing both the first and last deciles of wind speed is about two weeks. In week-2, the upper decile forecasts appear to be generally more skillful than the lower counterparts over southern Europe. In contrast, the lower decile forecasts appear to be more skillful than upper decile

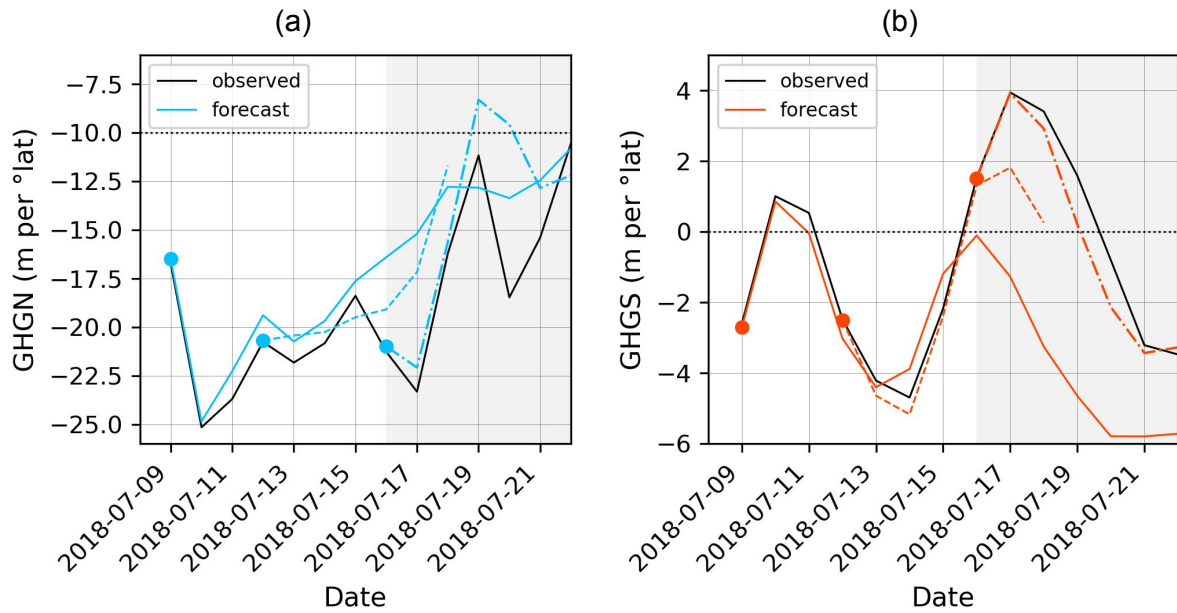


FIG. 8. Illustration of the daily evolution of atmospheric blocking indices over Europe. (a) GHGN and (b) GHGS. The reader is referred to Appendix a to learn about the computation of indices. The forecast indices shown in blue and orange lines indicate the ensemble mean of the indices. The solid, dashed, and dash-dotted lines correspond to the forecasts initialized on 9 July, 12 July, and 16 July, respectively (also shown in circles). The black curves correspond to the observed indices computed using ERA5 reanalysis. The grey-shaded region in (a) and (b) corresponds to the week between 16 and 22 July 2018. The values below the horizontal dotted line in (a) and above the horizontal dotted line in (b) indicate blocking.

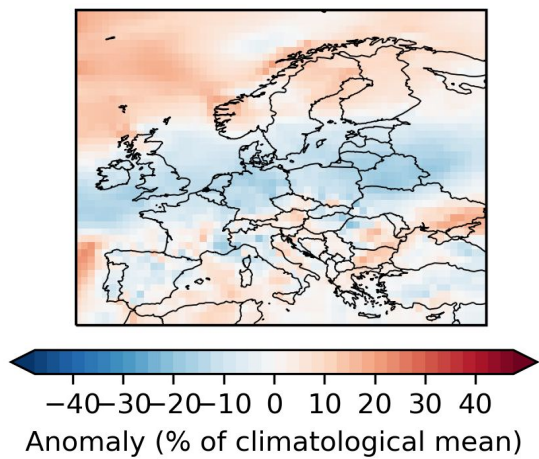


FIG. 9. Redundancy statistical prediction of 100-m wind speed anomaly for the week between 16 and 22 July 2018 over Europe using the observed Z500 anomaly over Euro-Atlantic for the same week (right panel in Figure 4) as the predictor.

forecasts over western Russia and to the northwest of the United Kingdom. This bias may either be attributed to the fact that the synoptic conditions which lead to strong winds over Europe are generally better predicted than those that

lead to weak winds (e.g., Grams et al. 2017; Zubiate et al. 2017; van der Wiel et al. 2019; Bloomfield et al. 2019; Cortesi et al. 2019; Garrido-Perez et al. 2020; Büeler et al. 2021; Cortesi et al. 2021) or that the period considered in here (i.e., boreal summer between 2016 and 2021) may have been more favorable for the prediction of one extreme with respect to the other due to favorable large-scale conditions (e.g. Jung et al. 2011). This possible bias noticeable in week-2 needs to be investigated further by computing the Brier Skill Score using the data spread over a longer period (for instance, using the re-forecasts). Beyond two weeks, there is virtually no difference in BSS between upper and lower decile dynamical predictions. Nonetheless, the upper decile hybrid predictions beyond two weeks show positive BSS over Germany. This may be due to the ability of the spatial patterns of variability of the statistical downscaling method to better reconstruct strong winds over this continental region. This sensitivity would need to be verified by repeating this exercise using the re-forecasts and by testing different predictor domains for reconstructing the 100-m wind speed. These questions will be explored in a future study.

To conclude, wind droughts are important weather phenomena having a significant impact on the energy sector. In this study, we investigated the wind drought which took



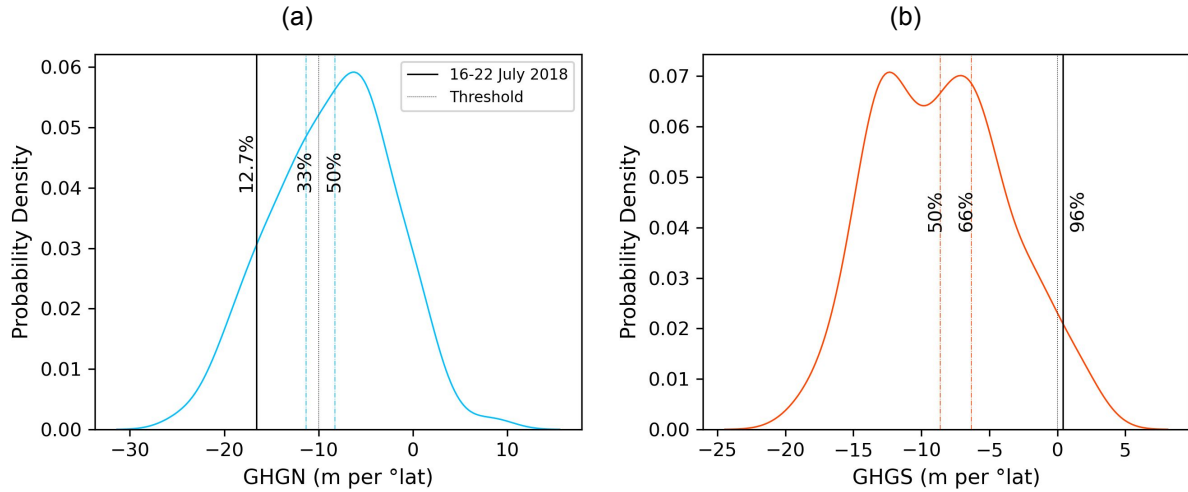


FIG. 10. Illustration of the probability density functions (PDFs) of atmospheric blocking indices over Europe for the boreal summer between 1999 and 2015 (i.e., training period used for statistical downscaling). (a) GHGN and (b) GHGS. The indices are computed for weekly mean values of Z500. The reader is referred to Appendix a to learn about the computation of indices. The PDFs are computed as kernel-density estimates (Gaussian kernel). The blue and orange dash-dotted vertical lines correspond to specific quantiles as indicated in the figure. The values to the left of the threshold (i.e., grey dotted line) in (a) and to the right of the threshold in (b) indicate blocking. The black solid lines in (a) and (b) indicate the values for the week between 16 and 22 July 2018.

place during the week between 16 and 22 July 2018 when wind speeds lower than 30% of the climatological normal were observed in and around the North Sea - the region with the highest density of wind farms in Europe. An atmospheric blocking event over eastern Scandinavia and the Atlantic Ocean was identified as causing the wind drought. Due to persistent difficulties of numerical weather prediction models in forecasting the onset and decay of atmospheric blocking events (e.g., Tibaldi and Molteni 1990; Ferranti et al. 2015; Matsueda and Palmer 2018; Büeler et al. 2021; Lupo 2021; Davini et al. 2021; Kleiner et al. 2021; Cortesi et al. 2021), the dynamical predictions show high uncertainty in anticipating wind droughts beyond a week. Similar to dynamical predictions, the hybrid predictions of 100-m wind speed obtained using redundancy analysis, although more skillful than the dynamical counterparts, equally fail to forecast the wind drought. Further analysis of the case helped us to understand the limitations of the statistical downscaling method in its current formulation in forecasting extreme events. Finally, the comparison of the Brier Skill Score between dynamical predictions of upper and lower decile wind speed over Europe suggests a possible bias of the ECMWF model in more skillfully predicting one extreme over the other.

*Acknowledgments.* Naveen Goutham would like to thank ANRT (Association Nationale de la Recherche et de la Technologie) for the CIFRE (Convention Industrielle de formation par la recherche) fellowship for his PhD.

This work contributes to the Energy4Climate Interdisciplinary Centre (E4C) of Institut Polytechnique de Paris and Ecole des Ponts ParisTech, supported by 3rd Programme d'Investissements d'Avenir [ANR-18-EUR-0006-02].

The authors acknowledge the data center ESPRI (Ensemble de services pour la recherche à l'Institut Pierre-Simon Laplace) for their help in storing and accessing the data.

*Data availability statement.* The archived ECMWF extended-range forecasts and re-forecasts are published under Creative Commons Attribution 4.0 International (CC BY 4.0). However, a data access fee may be applicable. For further information, kindly refer to <https://www.ecmwf.int/>. The ECMWF ERA5 reanalysis data is publicly available and can be accessed through the Climate Data Store of the Copernicus Climate Change Services upon registration.

## APPENDIX

### a. The atmospheric blocking indices over Europe

The geopotential height gradients in the north (GHGN) and the south (GHGS) at any given longitude are computed using a slightly modified version of the indices proposed by LejenÅs and Økland (1983) and Tibaldi and Molteni (1990). They are given below:

$$\text{GHGN} = \frac{Z500(\phi_n) - Z500(\phi_o)}{(\phi_n - \phi_o)} \quad (\text{A1})$$

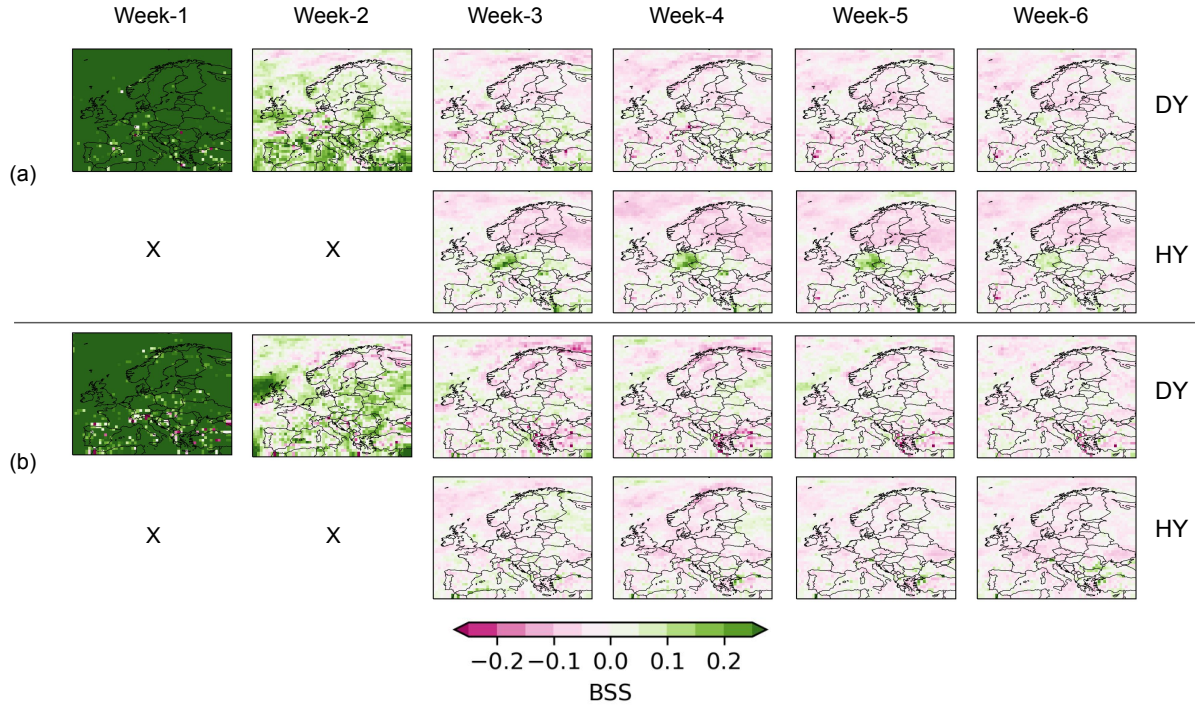


FIG. 11. Forecast quality assessment across the European domain, showing the weekly evolution of the Brier Skill Score (BSS) for dynamical (DY) and hybrid (HY) predictions. The BSS is computed for weekly mean 100-m wind speed forecasts of the upper (a) and lower deciles (b). The forecasts considered here are for the boreal summer (i.e. June, July, and August) between 2016 and 2021. Lead week-1 corresponds to the weekly average between days 0-6, week-1 between days 7-13, and so on. Positive BSS indicates that the forecasts are skillful relative to climatology.

and

$$\text{GHGS} = \frac{Z500(\phi_o) - Z500(\phi_s)}{(\phi_o - \phi_s)}. \quad (\text{A2})$$

Here,  $\phi$  denotes latitude where  $\phi_n = 80^\circ \text{ N}$ ,  $\phi_o = 65^\circ \text{ N}$ , and  $\phi_s = 45^\circ \text{ N}$ . We first compute GHGN and GHGS for daily mean values of Z500 at each of the longitudes over the domain between  $4.5^\circ \text{ E}$  and  $40^\circ \text{ E}$ . We then compute the zonal mean of GHGN and GHGS, and consider that the domain is blocked if either GHGS is  $> 0 \text{ m } ^\circ\text{lat}^{-1}$  or GHGN  $< -10 \text{ m } ^\circ\text{lat}^{-1}$ . We compute these indices for both the ensemble forecasts of Z500 and the corresponding observations (i.e., ERA5 reanalysis).

#### b. The Brier Skill Score

Although several scores exist to measure the accuracy of probabilistic forecasts of dichotomous events (i.e., yes/no events) (e.g., Jolliffe and Stephenson 2003; Wilks 2019), the most popular of them is the Brier Score (Brier 1950; Murphy 1986). The Brier Score (BS) is given by

$$\text{BS} = \frac{1}{n} \sum_{k=1}^n (y_k - o_k)^2, \quad (\text{A3})$$

where  $n$  is the number of forecast-observation pairs,  $y_k$  is the forecast probability,  $o_k$  is the observed probability ( $o_k$  is equal to 1 if the event occurs, or 0 otherwise). The relative value of forecasts with respect to climatology is quantified using the Brier Skill Score (BSS), given by  $\text{BSS} = 1 - (\text{BS}_{\text{forecast}} / \text{BS}_{\text{climatology}})$ . The forecast can be either dynamical or hybrid. Skillful forecasts must have positive BSS. The BSS in this work is computed for weekly mean 100-m wind speed forecasts of upper and lower deciles. The forecasts considered here are for the boreal summer between 2016 and 2021.

#### References

- Alonzo, B., 2018: Seasonal forecasting of wind energy resource and production in France and associated risk. Ph.D. thesis, Université Paris-Saclay.
- Alonzo, B., H.-K. Ringkjøb, B. Jourdie, P. Drobinski, R. Plougonven, and P. Tankov, 2017: Modelling the variability of the wind energy resource on monthly and seasonal timescales. *Renewable energy*, **113**, 1434–1446.
- Baldwin, M. P., D. B. Stephenson, D. W. J. Thompson, T. J. Dunkerton, A. J. Charlton, and A. O'Neill, 2003: Stratospheric Memory and Skill of Extended-Range Weather Forecasts. *Science*, **301** (5633), 636–640, <https://doi.org/10.1126/science.1087143>, publisher: American Association for the Advancement of Science Section: Report.

- Bell, B., H. Hersbach, A. Simmons, P. Berrisford, P. Dahlgren, A. Horányi, J. Muñoz-Sabater, J. Nicolas, R. Radu, D. Schepers, C. Soci, S. Villaume, J.-R. Bidlot, L. Haimberger, J. Woollen, C. Buontempo, and J.-N. Thépaut, 2021: The ERA5 global reanalysis: Preliminary extension to 1950. *Quarterly Journal of the Royal Meteorological Society*, **147** (741), 4186–4227, <https://doi.org/10.1002/qj.4174>.
- Bloomfield, H. C., D. J. Brayshaw, and A. J. Charlton-Perez, 2019: Characterizing the winter meteorological drivers of the European electricity system using targeted circulation types. *Meteorological Applications*, **27** (1), e1858.
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. 3.
- Brune, S., J. D. Keller, and S. Wahl, 2021: Evaluation of wind speed estimates in reanalyses for wind energy applications. *Advances in Science and Research*, **18**, 115 – 126, <https://doi.org/10.5194/asr-18-115-2021>.
- Buizza, R., 2019: Ensemble generation: The TIGGE and S2S ensembles. *Sub-Seasonal to Seasonal Prediction: The Gap between Weather and Climate Forecasting*, A. Robertson, and F. Vitart, Eds., Elsevier, chap. 13, 261–303.
- Buizza, R., and M. Leutbecher, 2015: The forecast skill horizon. *Quarterly Journal of the Royal Meteorological Society*, **141** (693), 3366–3382, <https://doi.org/10.1002/qj.2619>.
- Buizza, R., M. Leutbecher, and L. Isaksen, 2008: Potential use of an ensemble of analyses in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, **134** (637), 2051–2066.
- Buizza, R., M. Milleer, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, **125** (560), 2887–2908, <https://doi.org/10.1002/qj.49712556006>.
- Burton, T., N. Jenkins, D. Sharpe, and E. Bossanyi, 2011: *Wind energy handbook*. John Wiley & Sons.
- Büeler, D., L. Ferranti, L. Magnusson, J. F. Quinting, and C. M. Grams, 2021: Year-round sub-seasonal forecast skill for Atlantic–European weather regimes. *Quarterly Journal of the Royal Meteorological Society*, **147** (741), 4283–4309, <https://doi.org/10.1002/qj.4178>.
- Coelho, C. A. S., B. Brown, L. Wilson, M. Mittermaier, and B. Casati, 2019: Forecast verification for S2S timescales. *Sub-Seasonal to Seasonal Prediction: The Gap between Weather and Climate Forecasting*, A. Robertson, and F. Vitart, Eds., Elsevier, 337–361, section: 16.
- Cortesi, N., V. Torralba, N. González-Reviriego, A. Soret, and F. J. Doblas-Reyes, 2019: Characterization of European wind speed variability using weather regimes. *Climate Dynamics*, **53** (7-8), 4961–4976, <https://doi.org/10.1007/s00382-019-04839-5>.
- Cortesi, N., V. Torralba, L. Lledó, A. Manrique-Suñén, N. Gonzalez-Reviriego, A. Soret, and F. J. Doblas-Reyes, 2021: Yearly evolution of Euro-Atlantic weather regimes and of their sub-seasonal predictability. *Climate Dynamics*, **56** (11-12), 3933–3964, <https://doi.org/10.1007/s00382-021-05679-y>.
- Davini, P., A. Weisheimer, M. Balmaseda, S. J. Johnson, F. Molteni, C. D. Roberts, R. Senan, and T. N. Stockdale, 2021: The representation of winter northern hemisphere atmospheric blocking in ECMWF seasonal prediction systems. *Quarterly Journal of the Royal Meteorological Society*, **147** (735), 1344–1363, <https://doi.org/10.1002/qj.3974>.
- Domeisen, D. I. V., A. H. Butler, A. J. Charlton-Perez, B. Ayarzagüena, M. P. Baldwin, E. Dunn-Sigouin, J. C. Furtado, C. I. Garfinkel, P. Hitchcock, A. Y. Karpechko, H. Kim, J. Knight, A. L. Lang, E.-P. Lim, A. Marshall, G. Roff, C. Schwartz, I. R. Simpson, S.-W. Son, and M. Taguchi, 2020: The role of the stratosphere in subseasonal to seasonal prediction: 2. predictability arising from stratosphere-troposphere coupling. *J. Geophys. Res.: Atmospheres*, **125** (2), e2019JD030923, <https://doi.org/10.1029/2019JD030923>.
- Dörenkämper, M., B. T. Olsen, B. Witha, A. N. Hahmann, N. N. Davis, J. Barcons, Y. Ezber, E. García-Bustamante, J. F. González-Rouco, J. Navarro, M. Sastre-Marugán, T. Sile, W. Trei, M. Žagar, J. Badger, J. Gottschall, J. Sanz Rodrigo, and J. Mann, 2020: The making of the new European wind atlas – part 2: Production and evaluation. *Geoscientific Model Development*, **13** (10), 5079–5102, <https://doi.org/10.5194/gmd-13-5079-2020>.
- Ferranti, L., S. Corti, and M. Janousek, 2015: Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, **141** (688), 916–924, <https://doi.org/10.1002/qj.2411>.
- Fu, X., B. Wang, D. E. Waliser, and L. Tao, 2007: Impact of Atmosphere–Ocean Coupling on the Predictability of Monsoon Intraseasonal Oscillations. *Journal of the Atmospheric Sciences*, **64** (1), 157–174, <https://doi.org/10.1175/JAS3830.1>.
- Förster, H., S. Gores, C. Nissen, A. Siemons, N. Renders, S. Dael, M. Sporer, and M. Tomescu, 2021: Trends and projections in Europe 2021. Tech. rep., Copenhagen, Denmark. URL <https://www.eea.europa.eu/publications/trends-and-projections-in-europe-2021>.
- Gallo Cassarino, T., E. Sharp, and M. Barrett, 2018: The impact of social and weather drivers on the historical electricity demand in Europe. *Applied Energy*, **229**, 176–185, <https://doi.org/10.1016/j.apenergy.2018.07.108>.
- Garrido-Perez, J. M., C. Ordóñez, D. Barriopedro, R. García-Herrera, and D. Paredes, 2020: Impact of weather regimes on wind power variability in western Europe. *Applied Energy*, **264**, 114731, <https://doi.org/10.1016/j.apenergy.2020.114731>.
- Goutham, N., R. Plougonven, H. Omrani, S. Parey, P. Tankov, A. Tantet, P. Hitchcock, and P. Drobinski, 2022: How skillful are the European sub-seasonal predictions of wind speed and surface temperature? *Monthly Weather Review*, <https://doi.org/10.1175/MWR-D-21-0207.1>.
- Goutham, N., R. Plougonven, H. Omrani, A. Tantet, S. Parey, P. Tankov, P. Hitchcock, and P. Drobinski, Forthcoming: Statistical downscaling to improve the sub-seasonal predictions of energy-relevant surface variables. *Monthly Weather Review*, <https://doi.org/10.1175/MWR-D-22-0170.1>.
- Goutham, N., B. Alonzo, A. Dupré, R. Plougonven, R. Doctors, L. Liao, M. Mougeot, A. Fischer, and P. Drobinski, 2021: Using machine-learning methods to improve surface wind speed from the outputs of a numerical weather prediction model. *Boundary-Layer Meteorology*, **179** (1), 133–161, <https://doi.org/10.1007/s10546-020-00586-x>.
- Grams, C. M., R. Beerli, S. Pfenninger, I. Staffell, and H. Wernli, 2017: Balancing Europe’s wind-power output through spatial deployment informed by weather regimes. *Nature Climate Change*, **7** (8), 557–562, <https://doi.org/10.1038/nclimate3338>.

- Hersbach, H., B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. D. Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. d. Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut, 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146** (730), 1999–2049, <https://doi.org/10.1002/qj.3803>.
- IEA, 2020: European Union 2020. Tech. rep., Paris, France. URL <https://www.iea.org/reports/european-union-2020>.
- IEA, 2021: Net Zero by 2050. Tech. rep., Paris, France. URL <https://www.iea.org/reports/net-zero-by-2050>.
- Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud, 2010: Ensemble of data assimilations at ECMWF. (636), 45, <https://doi.org/10.21957/obke4k60>, publisher: ECMWF.
- Jiménez, P. A., J. V.-G. de Arellano, J. F. González-Rouco, J. Navarro, J. P. Montávez, E. García-Bustamante, and J. Dudhia, 2011: The effect of heat waves and drought on surface wind circulations in the Northeast of the Iberian Peninsula during the summer of 2003. *Journal of Climate*, **24** (20), 5416 – 5422, <https://doi.org/10.1175/2011JCLI4061.1>.
- Jolliffe, I., and D. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley, URL <https://books.google.com/books?id=cjS9kK81WBwC>.
- Jones, C., D. E. Waliser, K. M. Lau, and W. Stern, 2004a: Global Occurrences of Extreme Precipitation and the Madden–Julian Oscillation: Observations and Predictability. *Journal of Climate*, **17** (23), 4575–4589, <https://doi.org/10.1175/3238.1>.
- Jones, C., D. E. Waliser, K. M. Lau, and W. Stern, 2004b: The Madden–Julian Oscillation and Its Impact on Northern Hemisphere Weather Predictability. *Monthly Weather Review*, **132** (6), 1462–1471, [https://doi.org/10.1175/1520-0493\(2004\)132\(1462:TMOAI\)2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132(1462:TMOAI)2.0.CO;2), publisher: American Meteorological Society Section: Monthly Weather Review.
- Jourdir, B., 2020: Evaluation of ERA5, MERRA-2, COSMO-REA6, NEWA and AROME to simulate wind power production over France. *Advances in Science and Research*, Copernicus GmbH, Vol. 17, 63–77, <https://doi.org/10.5194/asr-17-63-2020>, iSSN: 1992-0628.
- Jung, T., F. Vitart, L. Ferranti, and J.-J. Morcrette, 2011: Origin and predictability of the extreme negative nao winter of 2009/10. *Geophysical Research Letters*, **38** (7), <https://doi.org/10.1029/2011GL046786>.
- Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, google-Books-ID: zx\_BakP2I5gC.
- Kleiner, N., P. W. Chan, L. Wang, D. Ma, and Z. Kuang, 2021: Effects of climate model mean-state bias on blocking underestimation. *Geophysical Research Letters*, **48** (13), e2021GL094129, <https://doi.org/10.1029/2021GL094129>.
- Koster, R. D., S. P. P. Mahanama, T. J. Yamada, G. Balsamo, A. A. Berg, M. Boissier, P. A. Dirmeyer, F. J. Doblas-Reyes, G. Drewitt, C. T. Gordon, Z. Guo, J.-H. Jeong, W.-S. Lee, Z. Li, L. Luo, S. Malyshev, W. J. Merryfield, S. I. Seneviratne, T. Stanelle, B. J. J. M. v. d. Hurk, F. Vitart, and E. F. Wood, 2011: The Second Phase of the Global Land–Atmosphere Coupling Experiment: Soil Moisture Contributions to Subseasonal Forecast Skill. *Journal of Hydrometeorology*, **12** (5), 805–822, <https://doi.org/10.1175/2011JHM1365.1>.
- LejenÅs, H., and H. Økland, 1983: Characteristics of northern hemisphere blocking as determined from a long time series of observational data. *Tellus A: Dynamic Meteorology and Oceanography*, **35** (5), 350–362, <https://doi.org/10.3402/tellusa.v35i5.11446>.
- Leutbecher, M., 2005: On ensemble prediction using singular vectors started from forecasts. (462), 11, <https://doi.org/10.21957/xuyeqttvx>, place: Shinfield Park, Reading Publisher: ECMWF.
- Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *Journal of computational physics*, **227** (7), 3515–3539.
- Leutbecher, M., P. Ollinaho, S.-J. Lock, S. Lang, P. Bechtold, A. Beljaars, A. Bozzo, R. Forbes, T. Haiden, R. Hogan, and I. Sandu, 2016: Model uncertainty representations in the IFS. *ECMWF/WWRP Workshop: Model Uncertainty*, ECMWF, Reading, URL <https://www.ecmwf.int/node/16369>.
- Li, B., S. Basu, S. J. Watson, and H. W. J. Russchenberg, 2021: A brief climatology of Dunkelflaute events over and surrounding the North and Baltic Sea areas. *Energies*, **14** (20), <https://doi.org/10.3390/en14206508>.
- Lin, H., and Z. Wu, 2011: Contribution of the Autumn Tibetan Plateau Snow Cover to Seasonal Prediction of North American Winter Temperature. *Journal of Climate*, **24** (11), 2801–2813, <https://doi.org/10.1175/2010JCLI3889.1>.
- Lledó, L., O. Bellprat, F. J. Doblas-Reyes, and A. Soret, 2018: Investigating the effects of pacific sea surface temperatures on the wind drought of 2015 over the united states. *Journal of Geophysical Research: Atmospheres*, **123** (10), 4837–4849, <https://doi.org/10.1029/2017JD028019>.
- Lledó, L., and F. J. Doblas-Reyes, 2020: Predicting Daily Mean Wind Speed in Europe Weeks ahead from MJO Status. *Monthly Weather Review*, **148** (8), 3413–3426, <https://doi.org/10.1175/MWR-D-19-0328.1>.
- Lupo, A. R., 2021: Atmospheric blocking events: a review. *Annals of the New York Academy of Sciences*, **1504** (1), 5–24, <https://doi.org/10.1111/nyas.14557>.
- Manwell, J. F., J. G. McGowan, and A. L. Rogers, 2010: *Wind energy explained: theory, design and application*. John Wiley & Sons.
- Manzanas, R., J. Gutiérrez, J. Fernández, E. van Meijgaard, S. Calmanti, M. Magariño, A. Cofiño, and S. Herrera, 2018: Dynamical and statistical downscaling of seasonal temperature forecasts in europe: Added value for user applications. *Climate Services*, **9**, 44–56, <https://doi.org/10.1016/j.cliser.2017.06.004>, climate services in practice: what we learnt from EUPORIAS.
- Manzanas, R., J. M. Gutiérrez, J. Bhend, S. Hemri, F. J. Doblas-Reyes, V. Torralba, E. Penabad, and A. Brookshaw, 2019: Bias adjustment and ensemble recalibration methods for seasonal forecasting: a comprehensive intercomparison using the C3S dataset. *Climate Dynamics*, **53** (3-4), 1287–1305, <https://doi.org/10.1007/s00382-019-04640-4>.
- Matsueda, M., and T. N. Palmer, 2018: Estimates of flow-dependent predictability of wintertime Euro-Atlantic weather regimes in medium-range forecasts. *Quarterly Journal of the Royal Meteorological Society*, **144** (713), 1012–1027, <https://doi.org/10.1002/qj.3265>.



- Michelangeli, P.-A., R. Vautard, and B. Legras, 1995: Weather regimes: Recurrence and quasi stationarity. *Journal of Atmospheric Sciences*, **52** (8), 1237–1256, [https://doi.org/10.1175/1520-0469\(1995\)052\(1237:WRRASQ\)2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052(1237:WRRASQ)2.0.CO;2).
- Miller, D. E., and Z. Wang, 2019: Skillful seasonal prediction of Eurasian winter blocking and extreme temperature frequency. *Geophysical Research Letters*, **46** (20), 11 530–11 538, <https://doi.org/10.1029/2019GL085035>.
- Molina, M. O., C. Gutiérrez, and E. Sánchez, 2021: Comparison of ERA5 surface wind speed climatologies over Europe with observations from the HadISD dataset. *International Journal of Climatology*, **41** (10), 4864–4878, <https://doi.org/10.1002/joc.7103>.
- Murcia, J. P., M. J. Koivisto, G. Luzia, B. T. Olsen, A. N. Hahmann, P. E. Sørensen, and M. Als, 2022: Validation of European-scale simulated wind speed and wind generation time series. *Applied Energy*, **305**, 117 794, <https://doi.org/10.1016/j.apenergy.2021.117794>.
- Murphy, A. H., 1986: A New Decomposition of the Brier Score: Formulation and Interpretation. *Mon. Wea. Rev.*, **114** (12), 2671–2673, [https://doi.org/10.1175/1520-0493\(1986\)114\(2671:ANDOTB\)2.0.CO;2](https://doi.org/10.1175/1520-0493(1986)114(2671:ANDOTB)2.0.CO;2).
- NewScientist, 2018: Weird 'wind drought' means Britain's turbines are at a standstill. URL <https://www.newscientist.com/article/2174262-weird-wind-drought-means-britains-turbines-are-at-a-standstill/>.
- Ohlendorf, N., and W.-P. Schill, 2020: Frequency and duration of low-wind-power events in Germany. *Environmental Research Letters*, **15** (8), 084 045, <https://doi.org/10.1088/1748-9326/ab91e9>.
- Olauson, J., 2018: ERA5: The new champion of wind power modelling? *Renewable Energy*, **126**, 322–331, <https://doi.org/10.1016/j.renene.2018.03.056>.
- Orsolini, Y. J., R. Senan, G. Balsamo, F. J. Doblas-Reyes, F. Vitart, A. Weisheimer, A. Carrasco, and R. E. Benestad, 2013: Impact of snow initialization on sub-seasonal forecasts. *Climate Dynamics*, **41** (7), 1969–1982, <https://doi.org/10.1007/s00382-013-1782-0>.
- Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. (598), 42, <https://doi.org/10.21957/ps8gbwbdv>, publisher: ECMWF.
- Prodhomme, C., F. Doblas-Reyes, O. Bellprat, and E. Dutra, 2016: Impact of land-surface initialization on sub-seasonal to seasonal forecasts over Europe. *Climate Dynamics*, **47** (3), 919–935, <https://doi.org/10.1007/s00382-015-2879-4>.
- Ramon, J., L. Lledó, V. Torralba, A. Soret, and F. J. Doblas-Reyes, 2019: What global reanalysis best represents near-surface winds? *Quarterly Journal of the Royal Meteorological Society*, **145** (724), 3236–3251, <https://doi.org/10.1002/qj.3616>.
- Raoult, B., C. Bergeron, A. L. Alós, J.-N. Thépaut, and D. Dee, 2017: Climate service develops user-friendly data store. ECMWF, Shinfield Park, Reading, URL <https://www.ecmwf.int/en/newsletter/151/meteorology/climate-service-develops-user-friendly-data-store>, newsletter 151.
- Reinhold, B., 1987: Weather regimes: The challenge in extended-range forecasting. *Science*, **235** (4787), 437–441, <https://doi.org/10.1126/science.235.4787.437>.
- Rex, D. F., 1950: Blocking action in the middle troposphere and its effect upon regional climate. *Tellus*, **2** (4), 275–301, <https://doi.org/10.3402/tellusa.v2i4.8603>.
- Robertson, A., and F. Vitart, 2018: *Sub-seasonal to Seasonal Prediction: The Gap Between Weather and Climate Forecasting*. Elsevier.
- Schwartz, C., and C. I. Garfinkel, 2020: Troposphere-stratosphere coupling in subseasonal-to-seasonal models and its importance for a realistic extratropical response to the Madden-Julian Oscillation. *J. Geophys. Res.: Atmospheres*, **125** (10), e2019JD032 043, <https://doi.org/10.1029/2019JD032043>.
- Seo, E., M.-I. Lee, J.-H. Jeong, R. D. Koster, S. D. Schubert, H.-M. Kim, D. Kim, H.-S. Kang, H.-K. Kim, C. MacLachlan, and A. A. Scaife, 2019: Impact of soil moisture initialization on boreal summer subseasonal forecasts: mid-latitude surface air temperature and heat wave events. *Climate Dynamics*, **52** (3), 1695–1709, <https://doi.org/10.1007/s00382-018-4221-4>.
- Siegert, S., and D. B. Stephenson, 2019: Forecast recalibration and multimodel combination. *Sub-Seasonal to Seasonal Prediction: The Gap between Weather and Climate Forecasting*, A. Robertson, and F. Vitart, Eds., Elsevier, chap. 15, 321–336.
- Sobolowski, S., G. Gong, and M. Ting, 2010: Modeled Climate State and Dynamic Responses to Anomalous North American Snow Cover. *Journal of Climate*, **23** (3), 785–799, <https://doi.org/10.1175/2009JCLI3219.1>.
- Soret, A., V. Torralba, N. Cortesi, I. Christel, L. Palma, A. Manrique-Suñén, L. Lledó, N. González-Reviriego, and F. J. Doblas-Reyes, 2019: Sub-seasonal to seasonal climate predictions for wind energy forecasting. *Journal of Physics: Conference Series*, **1222** (1), 012 009, <https://doi.org/10.1088/1742-6596/1222/1/012009>.
- Subramanian, A. C., M. A. Balmaseda, L. Centurioni, R. Chattopadhyay, B. D. Cornuelle, C. DeMott, M. Flatau, Y. Fujii, D. Giglio, S. T. Gille, T. M. Hamill, H. Hendon, I. Hoteit, A. Kumar, J.-H. Lee, A. J. Lucas, A. Mahadevan, M. Matsueda, S. Nam, S. Paturi, S. G. Penny, A. Rydbeck, R. Sun, Y. Takaya, A. Tandon, R. E. Todd, F. Vitart, D. Yuan, and C. Zhang, 2019: Ocean observations to improve our understanding, modeling, and forecasting of subseasonal-to-seasonal variability. *Frontiers in Marine Science*, **6**, <https://doi.org/10.3389/fmars.2019.00427>.
- TheGuardian, 2018: UK summer 'wind drought' puts green revolution into reverse. URL <https://www.theguardian.com/environment/2018/aug/27/uk-summer-wind-drought-puts-green-revolution-into-reverse>.
- Tibaldi, S., and F. Molteni, 1990: On the operational predictability of blocking. *Tellus A*, **42** (3), 343–365, <https://doi.org/10.1034/j.1600-0870.1990.t01-2-00003.x>.
- Tippett, M. K., T. DelSole, S. J. Mason, and A. G. Barnston, 2008: Regression-Based Methods for Finding Coupled Patterns. *Journal of Climate*, **21** (17), 4384–4398, <https://doi.org/10.1175/2008JCLI2150.1>.
- Torralba, V., F. J. Doblas-Reyes, D. MacLeod, I. Christel, and M. Davis, 2017: Seasonal Climate Prediction: A New Source of Information for the Management of Wind Energy Resources. *Journal of Applied Meteorology and Climatology*, **56** (5), 1231–1247, <https://doi.org/10.1175/JAMC-D-16-0204.1>.
- Toth, Z., and R. Buizza, 2019: Chapter 2 - weather forecasting: What sets the forecast skill horizon? *Sub-Seasonal to Seasonal Prediction*, A. W. Robertson, and F. Vitart, Eds., Elsevier, 17–45, <https://doi.org/10.1016/B978-0-12-811714-9.00002-4>.
- van den Hurk, B., F. Doblas-Reyes, G. Balsamo, R. D. Koster, S. I. Seneviratne, and H. Camargo, 2012: Soil moisture effects

- on seasonal temperature and precipitation forecast scores in Europe. *Climate Dynamics*, **38** (1), 349–362, <https://doi.org/10.1007/s00382-010-0956-2>.
- van der Wiel, K., H. C. Bloomfield, R. W. Lee, L. P. Stoop, R. Blackport, J. A. Screen, and F. M. Selten, 2019: The influence of weather regimes on European renewable energy production and demand. *Environmental Research Letters*, **14** (9), 094 010.
- Vautard, R., 1990: Multiple weather regimes over the North Atlantic: Analysis of precursors and successors. *Monthly Weather Review*, **118** (10), 2056 – 2081, [https://doi.org/10.1175/1520-0493\(1990\)118\(2056:MWROTN\)2.0.CO;2](https://doi.org/10.1175/1520-0493(1990)118(2056:MWROTN)2.0.CO;2).
- Vitart, F., A. W. Robertson, and D. L. Anderson, 2012: Subseasonal to Seasonal Prediction project: Bridging the gap between weather and climate. *Bulletin of the World Meteorological Organization*, **61** (2), 23.
- Vitart, F., R. Buizza, M. Alonso Balmaseda, G. Balsamo, J.-R. Bidlot, A. Bonet, M. Fuentes, A. Hofstadler, F. Molteni, and T. N. Palmer, 2008: The new VAREPS-monthly forecasting system: A first step towards seamless prediction. *Quarterly Journal of the Royal Meteorological Society*, **134** (636), 1789–1799.
- Vitart, F., C. Ardilouze, A. Bonet, A. Brookshaw, M. Chen, C. Codorean, M. Déqué, L. Ferranti, E. Fucile, M. Fuentes, H. Hendon, J. Hodgson, H.-S. Kang, A. Kumar, H. Lin, G. Liu, X. Liu, P. Malguzzi, I. Malas, M. Manoussakis, D. Mastrangelo, C. MacLachlan, P. McLean, A. Minami, R. Mladek, T. Nakazawa, S. Najm, Y. Nie, M. Rixen, A. W. Robertson, P. Ruti, C. Sun, Y. Takaya, M. Tolstykh, F. Venuti, D. Waliser, S. Woolnough, T. Wu, D.-J. Won, H. Xiao, R. Zaripov, and L. Zhang, 2017: The Subseasonal to Seasonal (S2S) Prediction project database. *Bulletin of the American Meteorological Society*, **98** (1), 163 – 173, <https://doi.org/10.1175/BAMS-D-16-0017.1>.
- Vitart, F., M. Alonso-Balmaseda, L. Ferranti, A. Benedetti, B. Balan-Sarajini, S. Tietsche, J. Yao, M. Janousek, G. Balsamo, M. Leutbecher, P. Bechtold, I. Polichtchouk, D. Richardson, T. Stockdale, and C. D. Roberts, 2019: Extended-range prediction. Tech. Rep. 854, ECMWF. <https://doi.org/10.21957/pdivp3t9m>.
- von Storch, H., F. Zwiers, and C. U. Press, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, URL [https://books.google.fr/books?id=\\\_VHxE26QvXgC](https://books.google.fr/books?id=\_VHxE26QvXgC).
- White, C. J., H. Carlsen, A. W. Robertson, R. J. T. Klein, J. K. Lazo, A. Kumar, F. Vitart, E. C. d. Perez, A. J. Ray, V. Murray, S. Bharwani, D. MacLeod, R. James, L. Fleming, A. P. Morse, B. Eggen, R. Graham, E. Kjellström, E. Becker, K. V. Pegion, N. J. Holbrook, D. McEvoy, M. Depledge, S. Perkins-Kirkpatrick, T. J. Brown, R. Street, L. Jones, T. A. Remenyi, I. Hodgson-Johnston, C. Buontempo, R. Lamb, H. Meinke, B. Arheimer, and S. E. Zebiak, 2017: Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications*, **24** (3), 315–325, <https://doi.org/10.1002/met.1654>.
- Wilks, D. S., 2014: Comparison of Probabilistic Statistical Forecast and Trend Adjustment Methods for North American Seasonal Temperatures. *Journal of Applied Meteorology and Climatology*, **53** (4), 935–949, <https://doi.org/10.1175/JAMC-D-13-0294.1>.
- Wilks, D. S., 2019: *Statistical methods in the atmospheric sciences*. 4th ed., Elsevier, Cambridge.
- WindEurope, 2022: Wind energy in Europe: 2021 statistics and the outlook for 2022–2026. WindEurope, Rue Belliard 40, B-1040 Brussels, URL <https://tinyurl.com/5bydexpd>.
- Woolnough, S. J., F. Vitart, and M. A. Balmaseda, 2007: The role of the ocean in the Madden–Julian Oscillation: Implications for MJO prediction. *Quarterly Journal of the Royal Meteorological Society*, **133** (622), 117–128, <https://doi.org/10.1002/qj.4>.
- Zheng, C., E. K.-M. Chang, H.-M. Kim, M. Zhang, and W. Wang, 2018: Impacts of the madden–julian oscillation on storm-track activity, surface air temperature, and precipitation over north america. *Journal of Climate*, **31** (15), 6113 – 6134, <https://doi.org/10.1175/JCLI-D-17-0534.1>.
- Zubiate, L., F. McDermott, C. Sweeney, and M. O’Malley, 2017: Spatial variability in winter NAO-wind speed relationships in western Europe linked to concomitant states of the East Atlantic and Scandinavian patterns: Variability of Winter Wind Speeds in Response to NAO, EA and SCA. *Quarterly Journal of the Royal Meteorological Society*, **143** (702), 552–562, <https://doi.org/10.1002/qj.2943>.
- Žúri, M., T. A. Huld, E. D. Dunlop, and H. A. Ossenbrink, 2007: Potential of solar electricity generation in the European Union member states and candidate countries. *Solar Energy*, **81** (10), 1295–1305, <https://doi.org/10.1016/j.solener.2006.12.007>.

# 5 Conclusions and Perspectives

Prediction is very difficult, especially if it is about the future.

— Niels Bohr

## 5.1 Conclusions

With a growing share of variable renewable power systems in the energy mix, the European energy sector is increasingly becoming weather sensitive. With Europe looking forward to a *climate neutral* future, the energy sector requires accurate predictions of essential climate variables beyond the deterministic range (i.e., two weeks) to facilitate smooth operation. In this regard, this Ph.D. served to provide information to the energy sector on sub-seasonal timescales, more informative than climatology, to aid in operational decision-making. This Ph.D. focused on two variables meaningful to the energy sector: the temperature at 2 m and wind speed at 100 m. The thesis was organized into three parts.

As an essential first step (Chapter 2), we quantified the skill of European sub-seasonal predictions of 100-m wind speed and 2-m temperature from the ECMWF using several quality metrics. Overall, depending on the geographical location and season, we observed temperature (skill horizon of about five weeks) to be generally more predictable than wind speed (skill horizon of about three weeks). This study provided a

*baseline* measure of the skill of sub-seasonal predictions from the ECMWF. Through this study, we hope to have strengthened the arguments in favor of the utility of sub-seasonal predictions in place of climatology for lead times of 2-5 weeks for energy-related applications.

As a second and significant step of the Ph.D. (Chapter 3), we developed a statistical downscaling methodology to infer surface temperature and wind speed using sub-seasonal predictions of the large-scale atmospheric circulation. We employed dimension reduction using historical data to capture the relationship between the large-scale atmospheric circulation and energy-relevant surface variables (i.e., temperature and wind speed) over Europe. We later exploited this relationship on sub-seasonal predictions of the large-scale atmospheric circulation, which are more reliable than surface variables, to infer temperature and wind speed predictions. Combining the dynamical (or direct) and the statistical (or indirect) predictions of surface variables, we obtain an ensemble twice as large, which we call hybrid predictions. Their skill was systematically assessed, demonstrating a significant improvement in the skill of sub-seasonal predictions of both temperature and wind speed. We attributed the improved accuracy of hybrid predictions to improved reliability and resolution. The implementation of this statistical downscaling methodology in the operational

decision-making value chain may bring additional value to the energy sector.

As a final part of the Ph.D., we studied episodes of wind drought over Europe due to their importance to the energy sector (Chapter 4). We specifically investigated the July 2018 wind drought due to its severity (weekly mean wind speed  $\sim 26\%$  below climatology over the North Sea). We analyzed the associated forecasts to determine the predictability horizon of the event. The wind drought of July 2018 was due to atmospheric blocking over eastern Scandinavia and the Atlantic Ocean. This case illustrated the difficulties in predicting wind droughts due to persistent challenges in forecasting the onset and decay of atmospheric blocking events. Finally, we highlighted the need for innovative methods to improve the prediction of atmospheric blocking.

## 5.2 Perspectives

Through this thesis, we demonstrated the value of dynamical and hybrid sub-seasonal predictions of 100-m wind speed and 2-m temperature over Europe for the energy sector. The energy mix of the European Union includes other sources of weather-sensitive renewable power such as hydropower (installed capacity of  $\sim 250$  GW in 2021 (TheEuropeanCommission, 2022a)) and solar photovoltaics (installed capacity of  $\sim 160$  GW in 2021 (TheEuropeanCommission, 2022b)). In this respect, the methodologies developed in this thesis are equally applicable to other energy-relevant climate variables such as solar radiation, precipitation, and streamflow. The methodologies may also be transferable to energy variables such as renewable energy production and electricity consumption.

In this thesis, we worked with predictions origi-

nating from a single model (i.e., ECMWF). Nevertheless, given the availability of sub-seasonal predictions from other models with different ways of accounting for errors in initial conditions and errors attributed to the imperfect knowledge of the atmosphere, the energy sector could make gains by working with multi-model predictions (e.g., Siegert and Stephenson, 2019). There exist several methods to combine multimodel ensembles (e.g. Stephenson et al., 2005; Coelho et al., 2006; Schepen et al., 2012; Wanders and Wood, 2016; Strazzo et al., 2019; Hemri et al., 2020; Brayshaw et al., 2020). The Postdoctoral research of Camille Le Coz, started in January 2022, contributes to this topic through the use of optimal transport distance (Peyré and Cuturi, 2019; Cumings-Menon and Shin, 2020). Her research is co-supervised by Rémy Flamary and Alexis Tantet at Laboratoire de Météorologie Dynamique (LMD), and I contribute to discussions on the choice of sub-seasonal forecasting systems and quality assessment methods. We have planned for collaborations to optimize the combination of dynamical and statistical predictions as discussed in Chapter 3.

The statistical downscaling methodology developed in Goutham et al. (Forthcoming) is based on a linear method. Although linear methods represent the relationship between the large-scale atmospheric circulation and the surface variables to a good approximation, they cannot fully explain the complex interactions between various components of the Earth system, and in particular non-linearities. The employment of non-linear methods such as neural networks in downscaling may further improve the skill horizon of sub-seasonal predictions. As an exploratory step in this direction, we started exploring the efficiency of convolutional neural networks, through the internship of Ganglin Tian at LMD which I co-supervised, in reconstruct-

ing sub-seasonal predictions of 100-m wind speed using predictions of the large-scale atmospheric circulation. The preliminary results from the internship are promising, and Ganglin will continue exploring non-linear methods to improve sub-seasonal predictions in his Ph.D. commencing late 2022 (co-supervised by Riwal Plougonven, Alexis Tantet, Anastase Charantonis, and Camille Le Coz).

### 5.2.1 Preliminary application

The sub-seasonal predictions from the ECMWF employed in this study forecast weather up to 46 days ahead. As the hybrid predictions developed in Goutham et al. (Forthcoming) remained skillful even at a lead time of six weeks for both wind speed and temperature, the statistical downscaling methodology can be equally extended to seasonal timescales. Having prior information about the expected renewable energy production and electricity consumption on seasonal timescales is increasingly becoming important for the energy sector. Additionally, we highlighted the possibility of tapping into other sources of information such as soil moisture, stratosphere, and sea surface temperature to further enhance the skill of sub-seasonal predictions (section 5b in Goutham et al. (Forthcoming)). In this regard, we carried out an exploratory study, through the internship of Omar Himych at EDF which I co-supervised along with Hiba Omrani, to forecast the cold spell of February 2018 on seasonal timescales. The French national electricity demand is highly thermo-sensitive. For example, with each degree Celsius drop in temperature below the threshold in winter, France requires about three additional nuclear reactors (~2.4 GW) to make up for the energy demand driven by space heating (RTE, 2019). This highlights the need for early detection of cold spells for smooth op-

erations of the energy sector. Since the cold spell of February 2018 was a result of Sudden Stratospheric Warming (e.g., Karpechko et al., 2018; Domeisen et al., 2020a,b; Kautz et al., 2020), we used geopotential height at 50 hPa (Z50) as the predictor.

To forecast the cold spell, instead of using monthly mean temperature, we used heating degree month (HDM) as the target variable. In addition to minimizing the loss of information, the choice of HDM as the target variable is driven by our intent to bridge the gap between meteorology and energy modeling communities. Although HDM is a very simplified way of representing the thermo-sensitivity of energy consumption, it is an interesting indicator as a first approximation. We computed HDM as the hourly difference between the threshold (in this case, threshold = 15°C) and the observed temperature, aggregated over the month. Accordingly, the higher the HDM, the colder is the month. Mathematically, the hourly temperature difference is calculated as

$$\Delta T_h = \begin{cases} T_{threshold} - T_h, & \text{if } T_h < T_{threshold} \\ 0, & \text{otherwise} \end{cases} \quad (5.1)$$

Then, HDM is conveniently computed as

$$\text{HDM} = \sum_{T_h \in \text{month}} \Delta T_h. \quad (5.2)$$

Furthermore, we retrained the redundancy analysis model between monthly anomalies of Z50 over the Euro-Atlantic and HDM over France using the ERA5 reanalysis. HDM, similar to temperature, shows a trend attributed to the warming climate. This trend was removed during pre-processing assuming a linear approximation (section 4c in Goutham et al. (Forthcoming)).

The training period considered was between 1979 and 2004. Upon training, we obtained new, paired spatial patterns of variability and the corresponding regression coefficients. Subsequently, we applied the regression coefficients on seasonal predictions of Z50 over the Euro-Atlantic from SEAS5 (Stockdale et al., 2018) of ECMWF to obtain seasonal statistical predictions of HDM over France. We added back the extrapolated trend to the statistical predictions of HDM. Figure 5.1 illustrates the observed and predicted HDM averaged over France for February 2018. The dynamical and statistical predictions in the figure correspond to those initialized on December 1, 2017. From the figure, it is conspicuous that the statistical predictions significantly outperform the dynamical counterparts in predicting the cold spell of February 2018, two months in advance.

Although seasonal prediction of the national average of HDM gives an estimate of the expected electricity demand, the energy sector requires gridded predictions of demand for operational purposes, e.g. to manage regional grids. Figure 5.2 compares the error in ensemble mean HDM between dynamical and statistical predictions. Overall, except for statistical predictions over the Alps, both the dynamical and statistical predictions under-estimate HDM. Nevertheless, it is conspicuous that the statistical predictions, with a mean error of  $-563^{\circ}\text{C-month}$  over the domain, significantly outperform dynamical predictions (mean error =  $-2146^{\circ}\text{C-month}$ ) in detecting the cold spell of February 2018. Although there remain several unanswered questions, the results from the internship of Omar are encouraging. This motivates further research into the employment of additional sources of predictability on sub-seasonal to seasonal timescales to detect extreme events such as cold spells and heat waves in advance to aid in the better management of

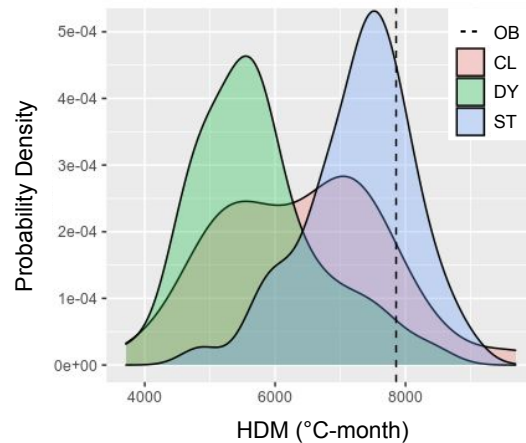


Figure 5.1: Illustration of the prediction of the cold spell of February 2018 averaged over France, showing the PDFs of dynamical (DY), statistical (ST), and climatological (CL) predictions. The dynamical predictions are initialized on 1 December 2017. The statistical predictions are produced using the monthly mean geopotential height at 50 hPa for February 2018 from the ensemble predictions initialized on 1 December 2017 as the predictor. The climatological prediction corresponds to the HDM of February from 1979 to 2017 computed using ERA5 reanalysis. The PDFs of ensemble members and climatology years are computed as kernel-density estimates assuming a Gaussian kernel. The dashed vertical line corresponds to the observed HDM (ERA5 reanalysis) in February 2018.

the future European power system.

## References

Brayshaw, D., P. Gonzalez, and F. Ziel, 2020: A new approach to subseasonal multi-model forecasting: Online prediction with expert advice. *EGU General Assembly Conference Abstracts*, <https://doi.org/10.5194/egusphere-egu2020-17663>.

Coelho, C. A. S., D. B. Stephenson, F. J.

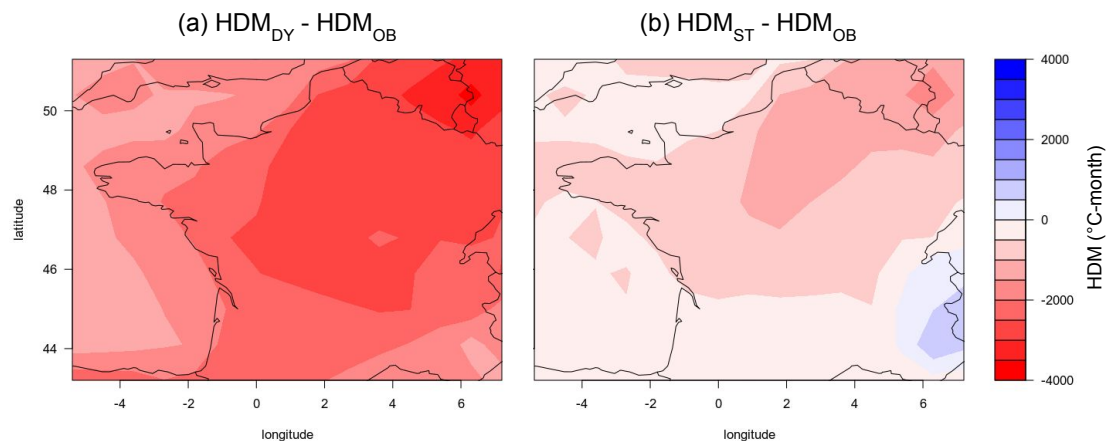


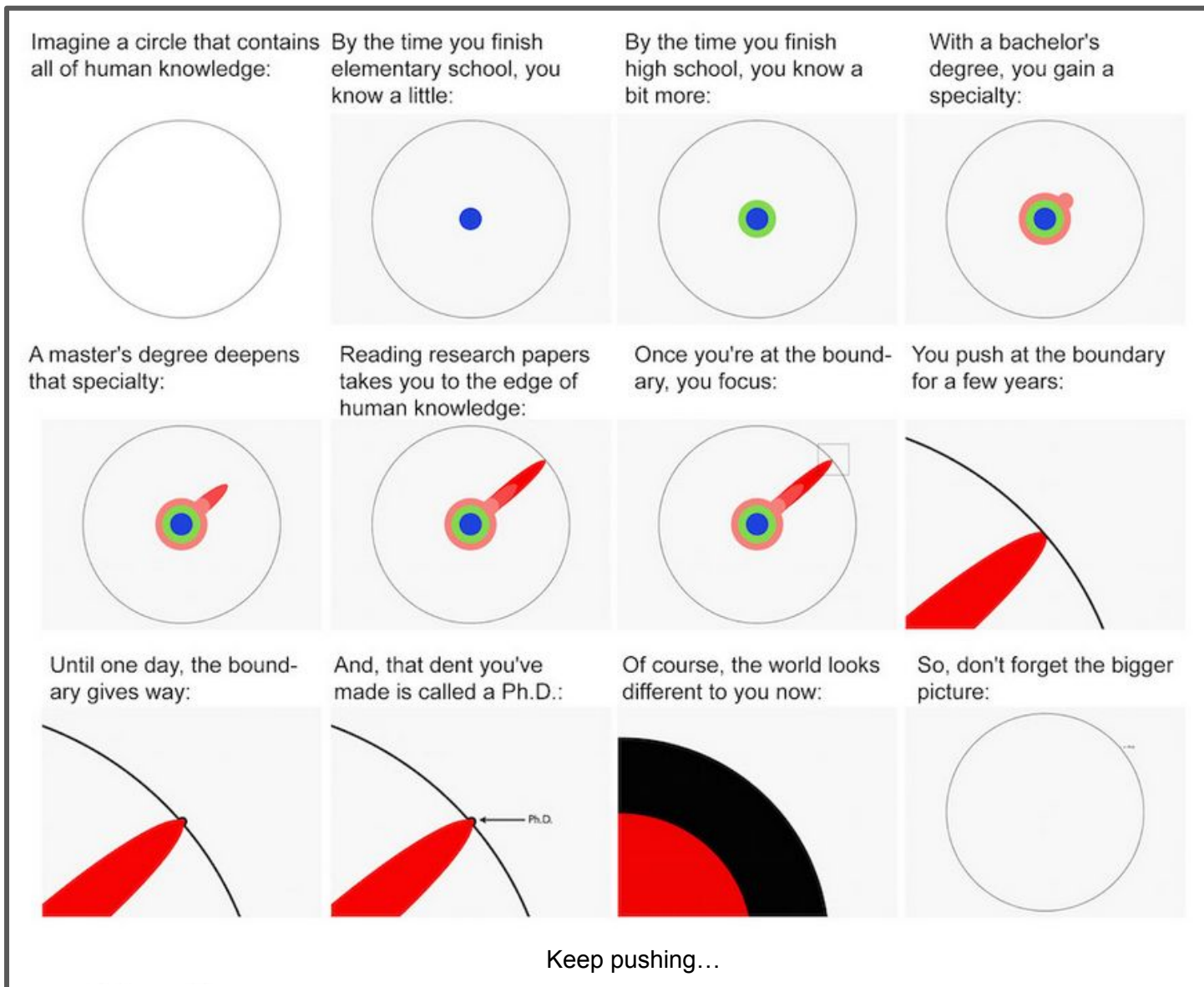
Figure 5.2: Comparison of error in HDM between (a) dynamical and (b) statistical predictions for February 2018 over France. The predictions shown here are initialized on December 1, 2017. The error is computed as the difference between the ensemble mean prediction and the observation (i.e., ERA5 reanalysis).

- Doblas-Reyes, M. Balmaseda, A. Guetter, and G. J. van Oldenborgh, 2006: A Bayesian approach for multi-model downscaling: Seasonal forecasting of regional rainfall and river flows in South America. *Meteor. Appl.*, **13** (01), 73, <https://doi.org/10.1017/S1350482705002045>.
- Cumings-Menon, R., and M. Shin, 2020: Probability forecast combination via entropy regularized Wasserstein Distance. *Entropy*, **22** (9), <https://doi.org/10.3390/e22090929>.
- Domeisen, D. I., A. H. Butler, A. J. Charlton-Perez, B. Ayarzagüena, M. P. Baldwin, E. Dunn-Sigouin, J. C. Furtado, C. I. Garfinkel, P. Hitchcock, A. Y. Karpechko, H. Kim, J. Knight, A. L. Lang, E.-P. Lim, A. Marshall, G. Roff, C. Schwartz, I. R. Simpson, S.-W. Son, and M. Taguchi, 2020a: The role of the stratosphere in Subseasonal to Seasonal prediction: 1. Predictability of the stratosphere. *Journal of Geophysical Research: Atmospheres*, **125** (2), e2019JD030920, <https://doi.org/10.1029/2019JD030920>.
- Domeisen, D. I. V., A. H. Butler, A. J. Charlton-Perez, B. Ayarzagüena, M. P. Baldwin, E. Dunn-Sigouin, J. C. Furtado, C. I. Garfinkel, P. Hitchcock, A. Y. Karpechko, H. Kim, J. Knight, A. L. Lang, E.-P. Lim, A. Marshall, G. Roff, C. Schwartz, I. R. Simpson, S.-W. Son, and M. Taguchi, 2020b: The role of the stratosphere in Subseasonal to Seasonal prediction: 2. Predictability arising from stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmospheres*, **125** (2), e2019JD030923, <https://doi.org/10.1029/2019JD030923>.
- Goutham, N., R. Plougonven, H. Omrani, A. Tantet, S. Parey, P. Tankov, P. Hitchcock, and P. Drobinski, Forthcoming: Statistical downscaling to improve the sub-seasonal predictions of energy-relevant surface variables. *Monthly Weather Review*, <https://doi.org/10.1175/MWR-D-22-0170.1>.
- Hemri, S., J. Bhend, M. A. Liniger, R. Manzananas, S. Siegert, D. B. Stephenson, J. M. Gutiérrez, A. Brookshaw, and F. J. Doblas-Reyes, 2020: How to create an operational



- multi-model of seasonal forecasts? *Climate Dynamics*, **55** (5), 1141–1157.
- Karpechko, A. Y., A. Charlton-Perez, M. Balmaseda, N. Tyrrell, and F. Vitart, 2018: Predicting Sudden Stratospheric Warming 2018 and its climate impacts with a multi-model ensemble. *Geophysical Research Letters*, **45** (24), 13,538–13,546, <https://doi.org/10.1029/2018GL081091>.
- Kautz, L.-A., I. Polichtchouk, T. Birner, H. Garny, and J. G. Pinto, 2020: Enhanced extended-range predictability of the 2018 late-winter Eurasian cold spell due to the stratosphere. *Quarterly Journal of the Royal Meteorological Society*, **146** (727), 1040–1055, <https://doi.org/10.1002/qj.3724>.
- Peyré, G., and M. Cuturi, 2019: *Computational Optimal Transport: With Applications to Data Science*. Foundations and trends in machine learning, Now Publishers, URL <https://books.google.fr/books?id=J0BiwgEACAAJ>.
- RTE, 2019: Electricity Report 2018. URL <https://bilan-electrique-2021.rte-france.com/wp-content/uploads/2019/02/BE-PDF-2018-1.pdf>.
- Schepen, A., Q. J. Wang, and D. E. Robertson, 2012: Combining the strengths of statistical and dynamical modeling approaches for forecasting Australian seasonal rainfall. *J. Geophys. Res.: Atmospheres*, **117** (D20), <https://doi.org/10.1029/2012JD018011>.
- Siegert, S., and D. B. Stephenson, 2019: Forecast recalibration and multimodel combination. *Sub-Seasonal to Seasonal Prediction: The Gap between Weather and Climate Forecasting*, A. Robertson, and F. Vitart, Eds., Elsevier, 321–336.
- Stephenson, D. B., C. A. S. Coelho, F. J. Doblas-Reyes, and M. Balmaseda, 2005: Forecast assimilation: a unified framework for the combination of multi-model weather and climate predictions. *Tellus A: Dynamic Meteorology and Oceanography*, **57** (3), 253–264, <https://doi.org/10.3402/tellusa.v57i3.14664>.
- Stockdale, T., M. Alonso-Balmaseda, S. Johnson, L. Ferranti, F. Molteni, L. Magnusson, S. Tietsche, F. Vitart, D. Decremmer, A. Weisheimer, C. D. Roberts, G. Balsamo, S. Keeley, K. Mogensen, H. Zuo, M. Mayer, and B. Monge-Sanz, 2018: SEAS5 and the future evolution of the long-range forecast system. (835), <https://doi.org/10.21957/z3e92di7y>.
- Strazzo, S., D. C. Collins, A. Schepen, Q. J. Wang, E. Becker, and L. Jia, 2019: Application of a hybrid statistical–dynamical system to seasonal prediction of North American temperature and precipitation. *Mon. Wea. Rev.*, **147** (2), 607 – 625, <https://doi.org/10.1175/MWR-D-18-0156.1>.
- TheEuropeanCommission, 2022a: Hydropower. URL [https://energy.ec.europa.eu/topics/renewable-energy/hydropower\\_en](https://energy.ec.europa.eu/topics/renewable-energy/hydropower_en).
- TheEuropeanCommission, 2022b: In focus: Solar energy – harnessing the power of the sun. URL <https://tinyurl.com/mr45mj55>.
- Wanders, N., and E. F. Wood, 2016: Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations. *Environmental Research Letters*, **11** (9), 094 007, <https://doi.org/10.1088/1748-9326/11/9/094007>.

# But, what is a PhD?



The above illustration is reproduced from "Matt Might, *The Illustrated Guide to a Ph.D.*, <https://matt.might.net/articles/phd-school-in-pictures/>", distributed under CC BY-NC 2.5 license.

## About the author

Upon graduating as a Mechanical Engineer in 2016 in India, Naveen moved to France to pursue a Master's in *Energy and Environment* at Ecole Polytechnique. The internship he carried out on improving wind speed from NWP model outputs in 2018 at Laboratoire de Météorologie Dynamique got him interested in weather forecasting. This thesis, started in January 2020 with quasi-zero knowledge about the subject, is a result of ~3 years of hard work while surfing a series of COVID waves. Naveen hopes you enjoy reading this work!



To know more: <https://gouthamnaveen.github.io/>

**Titre :** Prévisions météorologiques infra-saisonnnières pour le secteur de l'énergie en Europe : évaluation quantitative, amélioration et application

**Mots clés :** prévisions infra-saisonnnières, vitesse du vent, température, vérification des prévisions, prévisions statistiques, Europe

**Résumé :** Dans le cadre de la transition énergétique, la part des énergies renouvelables dans le mix énergétique est de plus en plus importante, rendant le système électrique plus sensible aux conditions météorologiques. En conséquence, le secteur de l'énergie est continuellement à la recherche de prévisions les plus précises possibles des variables climatiques sur un ensemble d'échelles de temps. Les prévisions météorologiques déterministes à court et moyen terme (de quelques minutes à deux semaines maximum) sont fiables, et leur utilisation opérationnelle dans le secteur de l'énergie est donc bien établie. Cependant, sur des échelles de temps infra-saisonnnières, c'est à dire au-delà de deux semaines et jusqu'à deux mois, les prévisions sont nécessairement probabilistes, et leur fiabilité reste limitée. Par conséquent, l'utilisation opérationnelle des prévisions infra-saisonnnières dans le secteur de l'énergie en est encore à ses débuts.

Disposer d'informations précises sur la production d'énergie renouvelable et la consommation d'électricité attendues sur des échelles de temps infra-saisonnnières peut apporter une vraie valeur ajoutée au secteur de l'énergie. De ce fait, l'objectif principal de cette thèse est d'évaluer en premier temps et d'améliorer ensuite les prévisions infra-saisonnnières par rapport à la climatologie, afin d'apporter des informations utiles et fiables au secteur de l'énergie. Nous nous concentrons dans ce travail sur la vitesse du vent à 100 m et la température à 2 m sur l'Europe.

Dans un premier temps, nous avons évalué les prévisions dynamiques infra-saisonnnières en termes de vent et de température afin de quantifier leurs performances telles qu'elles sont fournies par le modèle de prévision. Nous avons montré que les prévisions de la

température moyenne hebdomadaire sont plus fiables que la climatologie jusqu'à six semaines, et que celles de la vitesse du vent le sont jusqu'à trois semaines. Dans un deuxième temps, nous avons développé une technique de descente d'échelle statistique pour reconstruire des prévisions infra-saisonnnières de la vitesse du vent et de la température en utilisant les prévisions de variables climatiques de grande échelle. Pour ce faire, nous avons utilisé des données historiques observées pour estimer la relation entre la circulation atmosphérique à grande échelle et nos variables d'intérêt. Nous avons appliqué par la suite cette relation sur les prévisions infra-saisonnnières de la circulation à grande échelle, qui sont plus fiables que celles des variables de surface, pour en déduire des prévisions de nos variables d'intérêt. Cette méthode nous a permis de produire, à partir des prévisions infra-saisonnnières de la circulation à grande échelle, un nouvel ensemble de prévisions statistiques de température et de vent. Nous avons démontré que l'ensemble dit « hybride » combinant à la fois les nouvelles prévisions statistiques et les prévisions dynamiques de nos variables d'intérêt est plus fiable que les prévisions dynamiques seules. Pour la dernière partie de la thèse, nous avons développé une étude de cas sur les épisodes de faible vent en Europe, en raison de leur importance pour le secteur de l'énergie. Nous nous sommes intéressés à l'épisode de vents faibles de juillet 2018 et les prévisions associées. Pour cet événement, ni les prévisions dynamiques ni les prévisions statistiques n'ont réussi à le prévoir et ce en raison de la difficulté que les modèles de prévisions météorologiques ont à prévoir correctement les situations de blocage très souvent à l'origine de ces faibles vents.

**Title :** Sub-seasonal meteorological predictions for the European energy sector: quantitative assessment, improvement, and application

**Keywords :** sub-seasonal predictions, wind speed, temperature, forecast verification, statistical predictions, Europe

**Abstract :** Climate change has stimulated the energy sector, which is the largest emitter of global greenhouse gases (~40% in 2019), to transition to low-carbon energies. Europe, being one of the highest historical emitters of greenhouse gases, sits at the forefront of the energy transition. With a growing share of variable renewable power systems in the electricity mix on the one hand, and changing frequency and intensity of extreme events on the other, the weather-sensitive European energy sector is continuously on the lookout for accurate forecasts of essential climate variables on a continuum of timescales. The weather forecasts on short- to medium-range (i.e., from a few minutes ahead to at most two weeks) are reliable and essentially deterministic, and hence their operational use within the energy sector is well established. However, on timescales beyond two weeks and up to two months, i.e. in the sub-seasonal range, the predictions are necessarily probabilistic, and their reliability is far from that offered by short- and medium-range forecasts. Consequently, the operational use of sub-seasonal predictions within the energy sector is still in its infancy.

Having accurate information about the expected renewable energy production and electricity consumption on sub-seasonal timescales can help the energy sector in determining required reserve levels, scheduling maintenance, assessing and allocating risks attributed to extreme events, and estimating grid transmission capacity. In this regard, the main objective of this thesis is to provide more reliable information on sub-seasonal timescales, relative to climatology, to aid the energy sector in operational decision-making. We focus this research on 100-m wind speed and 2-m temperature over Europe. As an essential first step, we rigorously assess the skill of sub-seasonal dynamical predictions of these two variables to quantify their predictability limits as they are delivered in a given forecasting

system (the extended-range predictions of the European Centre for Medium-Range Weather Forecasts). We show that the weekly mean predictions of gridded temperature are more reliable than climatology for up to six weeks, and those of wind speed for up to three weeks. As a second step, we develop a statistical downscaling technique to reconstruct sub-seasonal predictions of wind speed and temperature using predictions of large-scale atmospheric circulation. We summarize the large-scale atmospheric state in a few indices by employing a dimension reduction methodology conditioned on wind speed and temperature over Europe. In other words, we use historical, observationally derived data to capture the relationship between the large-scale atmospheric circulation and our variables of interest (100 m wind speed and 2 m temperature). We then employ this relationship on sub-seasonal predictions of large-scale circulation, which are more reliable than surface variables, to deduce information about our variables of interest. This method allows us to produce, from a given ensemble of sub-seasonal predictions of large-scale circulation, a new ensemble of sub-seasonal predictions of our variables of interest. We demonstrate that the information thus extracted has value, as the hybrid ensemble combining both the dynamical and the statistical predictions of our variables of interest are more reliable than the dynamical predictions alone. As a final study, we investigate episodes of wind drought over Europe, because of their importance to the energy sector. A case study of the July 2018 episode of weak winds and the associated predictions, with and without our statistical downscaling methodology, illustrates the persistent difficulties of sub-seasonal predictions in predicting extreme events, in this case, due to the long-lasting challenge of forecasting blocking events.